AD-A020 004

SENSITIVITY ANALYSIS IN DISCRETE OPTIMIZATION

Michael Adrian Radke

California University

Prepared for:

Office of Naval Research

September 1975

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author)	2a REPORT SECURITY CLASSIFICATION
Western Management Science Institute University of California Los Angeles, CA 90024	Unclassified
	2b GROUP

3 REPORT TITLE

SENSITIVITY ANALYSIS IN DISCRETE OPTIMIZATION

4 DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Working Paper

5 AUTHOR(S) *(Last name first name initial)*

Radke, Michael A.

6 REPORT DATE	7a TOTAL NO OF PAGES	7b NO OF REFS
September, 1975	206	54

8a CONTRACT OR GRANT NO	9a ORIGINATOR'S REPORT NUMBER(S)
N00014-69-A-0200-4042	
b PROJECT NO	Working Paper No. 240
c	9b OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d	

10 AVAILABILITY/LIMITATION NOTICES

Available, request from: University of California Los Angeles, CA 90024

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY

13 ABSTRACT

Theory and methods are presented for analyzing sensitivity of the optimal value and optimal solution set to perturbations in problem data in nonlinear bounded optimization problems with discrete variables. Emphasis is given to studying behavior of the optimal value function. Theory is developed primarily for mixed integer programming (MIP) problems, where the domain is a subset of a Euclidean vector space.

Compared with continuous-variable mathematical programs, MIP problems demand much greater concern for data sensitivity and for model formulation. Continuous-variable LP (linear programming) problems do in fact produce sensible results in most applications because of the inherent continuity properties of the linear program. Unfortunately, the continuity properties that characterize LP models do not persist when integer restrictions are imposed. Evidently, great care must be given to an MIP formulation in order to achieve an acceptable formulation of a real world problem; sensitivity analysis is much needed as a means for determining problem behavior.

Conditions are developed that assure continuity at a point. Point-to-set mapping theory plays a central role: with reasonably weak assumptions on the objective function, a discontinuity in optimal value is possible only when the feasible region is discontinuous as a point-to-set map. When this map is "closed" the optimal value is a semicontinuous function of the data so that a "small"

DD FORM 1473   0101-807-6800

(continued)

change in data cannot create an incommensurate improvement in optimal value; this property always holds for bounded linear MIP problems. Continuity in the linear case is assured by a single post-optimal set-interiority condition. Formulation methods are prescribed that mitigate the bogey of discontinuity in some problems.

A general approach to determining sensitivity is described based on the construction of linear bounds. Several methods are given for discovering data perturbations that provide substantial improvements in the optimal value. With respect to data defining the feasible domain, this essentially amounts to locating points of discontinuity. Analysis of sensitivity to objective function data is especially straightforward since in virtually all practical discrete optimization problems the optimal value is continuous (and concave in most cases) with respect to such data.

Working Paper No. 240

(1) $B.S.$

SENSITIVITY ANALYSIS IN DISCRETE OPTIMIZATION

by

MICHAEL ADRIAN RADKE

September, 1975

WESTERN MANAGEMENT SCIENCE INSTITUTE

University of California, Los Angeles

WESTERN MANAGEMENT SCIENCE INSTITUTE

University of California, Los Angeles

Working Paper No. 240

SENSITIVITY ANALYSIS IN DISCRETE OPTIMIZATION*

by

Michael Adrian Radke


September, 1975

1 a

UNIVERSITY OF CALIFORNIA

Los Angeles

Sensitivity Analysis in Discrete Optimization

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Managemei ˙

by

Michael Adrian Radke

1975

1b

A

The dissertation of Michael Adrian Radke is approved, and it is acceptable in quality for publication on microfilm.

Donald Erlenkotter

Glenn W. Graves

Bruce L. Miller

Charles R. Scherer

Arthur M. Geoffrion, Committee Chairman

University of California, Los Angeles

1975

TABLE OF CONTENTS

iv

LIST OF FIGURES

ACKNOWLEDGMENTS

Finally, I congratulate my wife ███ and my offspring ██████,
████, and ██████ for their patience and understanding throughout this
research effort.

VITA

[PII Redacted] ▓▓▓▓▓▓▓▓▓▓--Born, ▓▓▓▓▓▓▓▓▓▓

1962--B.S., Physics/Mathematics/Music, Wisconsin State University

1962-1963--Instructor of Physics, Wisconsin State University

1963-1965--Graduate Teaching Assistant, Purdue University

1965--M.S., Applied Mathematics, Purdue University

1965-present--Member Professional Staff, TRW Systems, Redondo Beach, California

ABSTRACT OF THE DISSERTATION

Sensitivity Analysis in Discrete Optimization

by

Michael Adrian Radke

Doctor of Philosophy in Management

University of California, Los Angeles, 1975

Professor Arthur M. Geoffrion, Chairman

Theory and methods are presented for analyzing sensitivity of the
optimal value and optimal solution set to perturbations in problem data
in nonlinear bounded optimization problems with discrete variables.
Emphasj is given to studying behavior of the optimal value function.
Theory is developed primarily for mixed integer programming (MIP) prob-
lems, where the domain is a subset of a Euclidean vector space.

Compared with continuous-variable mathematical programs, MIP prob-
lems demand much greater concern for data sensitivity and for model
formulation. Continuous-variable LP (linear programming) problems
do in fact produce sensible results in most applications because of
the inherent continuity properties of the linear program. Unfortunate-
ly, the continuity properties that characterize LP models do not persist
when integer restrictions are imposed. Evidently, great care must be
given to an MIP formulation in order to achieve an acceptable formula-
tion of a real world problem; sensitivity analysis is much needed as
a means for determining problem behavior.

Conditions are developed that assure continuity at a point. Point-
to-set mapping theory plays a central role: with reasonably weak assump-
tions on the objective function, a discontinuity in optimal value is

possible only when the feasible region is discontinuous as a point-to-set map. When this map is "closed" the optimal value is a semicontinuous function of the data so that a "small" change in data cannot create an incommensurate improvement in optimal value; this property always holds for bounded linear MIP problems. Continuity in the linear case is assured by a single post-optimal set-interiority condition. Formulation methods are prescribed that mitigate the bogey of discontinuity in some problems.

A general approach to determining sensitivity is described based on the construction of linear bounds. Several methods are given for discovering data perturbations that provide substantial improvements in the optimal value. With respect to data defining the feasible domain, this essentially amounts to locating points of discontinuity. Analysis of sensitivity to objective function data is especially straightforward since in virtually all practical discrete optimization problems the optimal value is continuous (and concave in most cases) with respect to such data.

CHAPTER 1

SENSITIVITY ANALYSIS: A PERSPECTIVE

One of the most worrisome properties of mixed integer programming
(MIP) problems is the plaguing discontinuity of the optimal solution
value with respect to problem data, a characteristic that is compara-
tively infrequent (although present) in continuous-variable problems.
With minimization problems in particular, we are referring to the possi-
bility that a slight perturbation in the data defining the problem may
create an incommensurately large reduction in the optimal value. It
will be shown that this situation can occur only when the feasible
region exhibits a discontinuous behavior with respect to data changes,
provided the objective function is well-behaved. Discontinuous behavior
of the optimal value and optimal solution set is a model instability
that can jeopardize the successful managerial application of MIP models.
Sensitivity analysis in discrete optimization is dedicated to the iden-
tification, interpretation, and alleviation of such instabilities.

1.1 Motivation for Sensitivity Analysis

Currently, the study of sensitivity in mixed integer programming is
in its very early stages. This is understandable since until recently
algorithms for solving MIP problems were relatively inefficient and
further analysis would be an overburdening expense. Fortunately, the
state-of-the-art in MIP has evolved substantially in the past few years
and it is hoped that this treatise will mark a genesis of sensitivity
analysis practices and will spark further development in the analysis
of MIP sensitivity. A compendium of recent advances is given by
Geoffrion [12], and earlier advances are surveyed in Geoffrion and

1

Marsten [13].

Compared with continuous-variable mathematical programs MIP problems demand much greater concern for data sensitivity and for model formulation. The reason that (continuous-variable) LP problems do in fact produce sensible results in most applications is due to the inherent continuity properties of the linear program (see Chapter 6). That is to say, the analyst accepts his LP formulation of the problem as an adequate representation of the real world because he is confident that a small change in data will permit only a small change in the optimal solution value.

Unfortunately, the soothing continuity properties that characterize LP models do not persist when integer restrictions are imposed. Consequently, constraint formulations deemed as adequate in an LP model may not be acceptable in a mixed integer model in the sense that unrealistic discontinuities can occur. Evidently, greater care must be given to the formulation of an MIP problem in order to achieve an acceptable formulation of the real world.

Practical problems that are formulated as mathematical programs are seldom completely solved by the determination of a solution to the mathematical program. This is especially true for linear formulations because they so frequently express lenient approximations to the real problem. The problem data are the link between reality and the model -- seldom, in practical problems, are the data truly constant. Consequently, it is ordinarily advisable to perform a sensitivity analysis to determine the effect of data changes on the optimal solution and solution value. Based on sensitivity analysis results the analyst may want to alter the data slightly or perhaps improve the model formulation

so as to remove high sensitivities.

There are several reasons why data may not be constant and why, therefore, perturbations in the data must be investigated. Perhaps the most common is the fact that data is seldom known with certainty -- it is estimated. A data point may be an approximation based on statistical samples; it may be stochastic in nature. For instance, in capital budgeting, the return on an investment portfolio is a prediction based on past experience and expected economic influences.

Another common property of data is that it varies with time. There are many examples: demands in facility location, costs of food items in menu-planning, the availability of resources in resource allocation, etc. When such data are held fixed the resulting static model is meaningful only over a short period of time. It is common practice to approximate the dynamic nature of a system by a sequence of static models, solved in parametric fashion if possible (see Nauss [22]).

Sensitivity analysis can be useful in run design, e.g., in determining the direction and size of a data change for subsequent static optimizations. In general, the data may vary (slightly) with respect to parameters other than time, perhaps even with respect to variables explicit in the model.

Frequently, data are allowed to vary arbitrarily, but for a price. Budgets can always be stretched as long as lending institutions are in business; additional materials needed in manufacturing can usually be obtained, perhaps at a premium price, warehouse design capacity can be increased by accepting an alternative design with a higher construction/ operating cost; product demand requirements might be reduced but with a penalty to future business.

3

When such data are modeled as constants it is usually done under the presumption that small data perturbations will not induce incommensurately large changes in optimal value. This permits a simpler model and avoids having to gather the additional prices, costs, and/or utilities which would be needed to model data variations explicitly. The presumption is often a bad one for MIP models. That it is commonplace may possibly be due to a false sense of security carried over from LP modeling practices.

Another situation is rather embarrassing but it occurs nonetheless quite frequently. After solving the problem it may be discovered that a datum is wrong, perhaps a misplaced decimal point or a keypunch error. Less embarrassing is the case where new information indicates that the estimates of certain data values should be revised.

We have just identified basically four types of data that motivate exercising sensitivity analysis:

- data with uncertainty
- dynamic data
- subjective data
- wrong data.

Each of these data types may suggest slightly different motives for sensitivity analysis. In the case of uncertain data -- which result from predictions, physical measurements, statistical sampling, etc. -- the analyst is concerned first with stability; that is, some assurance is sought that small changes about the given data point will yield likewise small changes in optimal value. With regard to physical data which represent measurements, assurance of stability may be a sufficient objective; but depending on the nature of the data the analyst may

4

desire quantitative information on the variation of the optimal value and solution with respect to prescribed data variations.

In some cases, particularly with dynamic data whose fluctuations with time are predictable, the sensitivity analysis may be focused on directional changes in the data. Customer demands, for example, might be predicted to increase proportionately over time, in which case the sensitivity in optimal value can be expressed in terms of time, a scalar quantity.

Subjective data (managerial and operational specifications, guestimates, utilities, etc.) are commonly perturbable, particularly when a reasonable change in such data can be shown to result in a sizable improvement in the optimal value. A budgetary limitation might be relaxed, for instance, if the resulting reward favorably exceeds the cost of the expanded budget. The motive in this case appears to be to discover reasonable perturbations in the data that yield net improvements in optimal value (see Chapter 11).

In the situation where data are discovered to be in error after-the-fact, the subsequent response is typically a frantic attempt to reconstruct an optimal or near-optimal solution from the erroneously derived solution. Bounds on the deviation in optimal value (and solutions corresponding to the bounds) may possibly be the best that can be expected from a post-optimal analysis; the methods of sensitivity analysis can be of some assistance.

1.2 The Principal Issues

The nature and amount of sensitivity information desired for a particular problem depends mainly on the characteristics of the data,

5

especially the degree of flexibility of the data. Data that are flexible are allowed to vary substantially from their specified values, usually at some additional expense. A data point that is permitted to be changed by one or more percentage points can certainly be regarded as flexible. Common examples of flexible data are budgets, material supplies, design specifications such as capacity, production levels, etc. Flexible data can be viewed as model parameters whose assigned values may be reasonably perturbed if the net payoff for doing so is substantial. What exactly is a reasonable perturbation, of course, is dictated by the degree of data flexibility. Rigid data such as transport distances, system component weights, the volume of an available fuel tank, etc., obviously have little or no active role in a sensitivity analysis.

Given the problem and specified data values the first question perhaps is whether or not the optimal value is continuous with respect to the data. This is the question of stability; unstable behavior must be explained: either the model is an inaccurate representation of the real world or there is some physical justification for the behavior. Secondly, depending on flexibility of the data, the analysis may be directed toward discovery of reasonable data perturbations that yield favorable improvements in the optimal value.

Identification of such improving data changes and their corresponding optimal values provides the manager with a collection of alternatives that can be analyzed with regard to utilities and complex considerations that were not represented in the model. In addition, the sensitivity analysis should provide interpretive aids over and above the fact that a specific data change buys a specific improvement in

6

optimal value, such as rates of change, bounds on the amount of change, and identification of critical data and critical constraints.

This kind of sensitivity information is useful also to the analyst should he decide that the model must be reformulated to achieve greater realism. Instabilities (discontinuities) that are interpreted as due to oversimplification and approximation in the model give rise to the issue of how such modeling inadequacies may be alleviated. If a datum is known to be critical (in the sense that it is responsible for an unrealistic discontinuity) then its replacement by a variable may yield a satisfactory reformulation of the model. Similarly, critical constraints might be remodeled by relaxation with a violation penalty in the objective function. Such techniques for model formulation and reformulation are the topic of Chapter 7.

The purpose of a sensitivity analysis is to investigate one or more of the issues posed by questions such as:

(1) Can a small change in the data permit a relatively large change in the optimal value? If so, which data and which constraints are responsible?

(2) How much can data be varied without creating an incommensurate change in the optimal value and solution?

(3) What particular data perturbations permit improvements in the optimal value and how much improvement?

(4) What is the rate of change of the optimal value with respect to (componentwise or directional) data changes?

(5) In the case of parametrized data, what is the approximate behavior of the optimal value as a function of the data parameter?

7

Issues one and two are supremely subjective in nature, but primarily they are concerned with the existence of discrete jumps (discontinuities) in the optimal value that occur due to perturbations in the data: is the optimal value discontinuous at or very near to the given data point? If not, how much (and in what manner) can the data be varied without permitting any discontinuity in the optimal value? The second question is, of course, more difficult to answer than the first.

The third question asks for what is perhaps the most that can be expected from a sensitivity analysis: a collection of alternative data values and corresponding improvements in the optimal value. However, for certain restricted variations in data, e.g., variation of only one component of the data or of parametrized data, it can be practical to estimate the optimal value as a function of the scalar parameter. Thus, if $b + \theta r$ denotes the right side of a set of inequality constraints, the objective is to estimate the optimal value function $v(\theta)$ over a specified range of values for θ (this is question (5)).

Some insight into these issues of sensitivity analysis is provided by the illustrative observational analysis in the next section.

1.3 An Illustrative Analysis of a Simple Problem

$$\text{maximize } 200y_1 + 160y_2 + 60y_3 + 40y_4$$
$$\text{subject to: } 41y_1 + 39y_2 + 20y_3 + 10y_4 \leq 80$$
$$30y_1 + 25y_2 + 20y_3 + 15y_4 \leq 70 \tag{1.1}$$
$$y_i = 0 \text{ or } 1, \quad (i = 1,2,3,4) .$$

The two budgets will be referred to as b_1 and b_2 ($b_1 = 80$, $b_2 = 70$); let c denote the objective function coefficients and let v denote the optimal value of problem (1.1). For the given values of b and c,

8

$v = 360$ and $y^* = (1,1,0,0)$ is uniquely optimal.

We will analyze the sensitivity of v to changes in b and c with regard to questions (1) through (5) of the previous section. Because (1.1) is so simple the sensitivity information can be obtained by observation. To obtain similar information for more complex problems will generally require computational procedures that exploit theoretical continuity and sensitivity properties. This brief exercise is intended to exhibit the primary issues of a sensitivity analysis and to suggest concepts and procedures that generalize to more complex applications.

Consider the first question: can a small change in c or b (or both) cause a large improvement in v? It is clear that v is monotone nondecreasing with respect to increases in c. It is also easy to see that v cannot take a discrete jump in value due to a change in c because any solution optimal at c is feasible for any (bounded) perturbations in c. That is, v varies continuously with respect to changes in c. Furthermore, it is relatively easy to derive bounds on the variation in v with respect to changes in c (this topic is studied in Chapter 9).

Small changes in b on the other hand can give rise to discrete changes in v. If b_1 is decreased (from 80) by any amount however small the current optimal solution y^* becomes infeasible and v takes a discrete jump downward (to 300). However, we are mostly interested in data changes that might improve the optimal value (increases in b in this case). Can a small increase in b admit a solution with a much higher optimal value? Consider the increased budget $b = (85,75)$. Resolving the problem (by inspection in this simple example), $y^* = (1,1,0,0)$ is found to still be optimal so that v remains equal to 360 for increases in b of roughly 6% or less.

9

Evidently, for problem (1.1) the answer to question (1) is: no, a small change in data does not permit a relatively large improvement in the optimal value v. At this point the analyst either is satisfied that his problem is stable with respect to the data, or he is interested in knowing how much (further) the data can be varied without producing a violent change in v or a change in the optimal solution. The amount or type of sensitivity information desired by the analyst depends on the rigidity or flexibility of the data.

Suppose that the budgets b_1 and b_2 could at some per unit cost be increased by as much as 5% but under no circumstances could they be stretched by more than 5%. In this case the analyst is content with the answer to question (1) and the analysis is complete. On the other hand, suppose that the budgets may be stretched by as much as 20% (for a price). Now the analyst will want to continue to analyze. How much can b be stretched without causing a jump in v? This is question (2).

If b_1 is increased to 90, $\hat{y} = (1,1,0,1)$ becomes the (unique) optimal solution for a value of $v = 400$. For values of b_1 less than 90 (and greater than 80) $y^* = (1,1,0,0)$ and $v = 360$ remain optimal; in fact, they are optimal for all b such that $80 \leq b_1 < 90$ and $55 \leq b_2 < \infty$.

We have obtained a partial answer to question (3): an increase in b_1 of 10 units increases v by 40. In addition, we find that increasing b to (100,85) increases v by 60 ($v = 420$ with $y^* = (1,1,1,0)$); further increasing b to (110,100) yields $v = 460$ with $y^* = (1,1,1,1)$.

In this simple example it is clear that these are the only points of discontinuity that can result by allowing b to increase. In general, one would want to discover those points of discontinuity that arise from reasonable data perturbations and that provide a net improvement

10

in the objective value.

Consideration (4) -- the rate of change of v -- applies only to data points at which v is continuous. For instance, it makes sense to determine the rate of change of v with respect to objective function coefficients since (as we shall formally establish later) in any mixed integer program v is continuous with respect to such data. Because v as a function of c is given by $v(c) = c \cdot y^*$, we have for problem (1.1) that $dv/dc = y^*$ (i.e., $dv/dc_1 = 1$, $dv/dc_2 = 1$, $dv/dc_3 = 0$, $dv/dc_4 = 0$) at $c = (200,160,60,40)$ and $b = (80,70)$. If we consider parametric changes in c of the form $\gamma \cdot (1,1,1,1)$ where $\gamma \geq 0$, then $dv/d\gamma = 2$ at the given values of c and b.

Rates of change of v, or more accurately, directional derivatives, are of interest for two reasons. First, even if v varies continuously with respect to a (continuous) change in data, v can be highly sensitive to the data. This sensitivity is measured by a directional derivative. Secondly, we will see later that these derivatives, or estimates thereof, are useful in procedures for discovering points of discontinuity.

Data that influence the feasible region, such as b_1 and b_2 in (1.1), are responsible for any discontinuity in v and in the feasible region itself. For problem (1.1), let $\frac{dv}{db_1}\Big|^{+}$ denote the rate of change of v with respect to increases in b_1. Similarly, $\frac{dv}{db_1}\Big|^{-}$ is a left-hand derivative. At $b = (80,70)$ we have

$$\frac{dv}{db_1}\Big|^{-} = -\infty, \quad \frac{dv}{db_1}\Big|^{+} = 0$$

and at $b = (85,75)$, both derivatives are zero.

In pure integer problems these one-sided derivatives of v with respect to a right-side datum are always either zero or infinite in

11

value. We will find these derivatives to be more interesting in the mixed integer case.

With regard to question (5), one could plot v as a function of γ where the objective function coefficients are given by $c + \gamma c'$ and b is held fixed, as a function of β where the right sides are $b + \beta b'$ and c is held fixed, or as a function of θ where the objective coefficients and right-side parametrizations are $c + \theta c'$ and $b + \theta b'$, respectively. These functions -- $v(\gamma)$, $v(\beta)$, and $v(\theta)$ -- for problem (1.1) are plotted in Figures 1.1, 1.2, and 1.3. We have arbitrarily selected $c' = (-50, -40, 30, 25)$ and $b' = (30, 25)$; the parameters are permitted to vary from 0 to 1.

Notice that v is continuous as a function of γ but is discontinuous as a function of β. The function $v(\theta)$ is also discontinuous; $v(\theta)$ is the optimal value function of the problem:

maximize $(200-50\theta)y_1 + (160-40\theta)y_2 + (60+30\theta)y_3 + (40+20\theta)y_4$

$\quad$ s.t. $41y_1 + 39y_2 + 20y_3 + 10y_4 \le 80 + 30\theta$

$\qquad 30y_1 + 25y_2 + 30y_3 + 15y_4 \le 70 + 25\theta$

$\qquad y_i = 0$ or 1, $\quad (i = 1,2,3,4)$.

These plots were obtained by observation; methods for efficiently obtaining or approximating similar plots for more complex problems are studied in Chapters 9, 10, and 11.

An important objective in the development of computational procedures is to format the sensitivity information in a way that is meaningful and yet is easy to interpret. The sensitivity issues (1) through (5) of paragraph 1.2 were phrased to accommodate this objective. For parametric data two-dimensional plots are reasonably derivable and easy

12

to interpret. For nonparametric data functional relationships are un-
reasonably difficult to derive and to interpret; however, a collection
of alternate data points that yield improved solutions is reasonably
obtainable information.
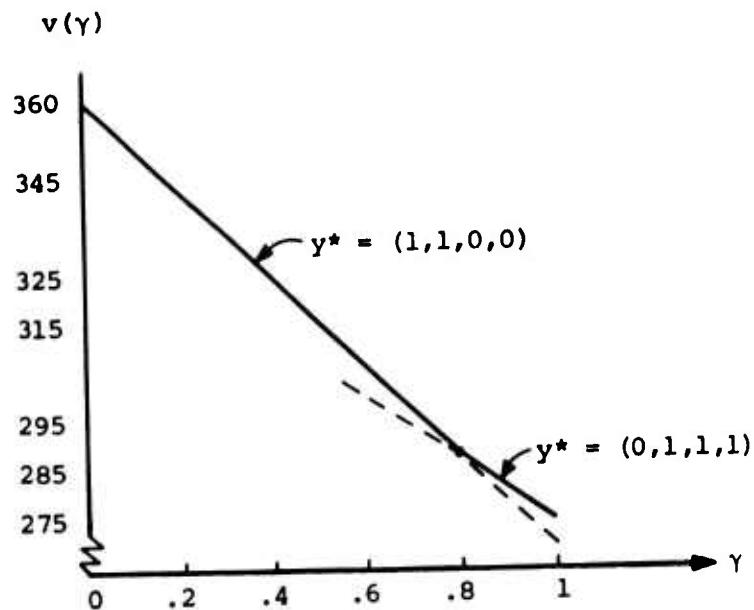
Figure 1.1.   Problem (1.1)--v as a Function of Parametric Costs
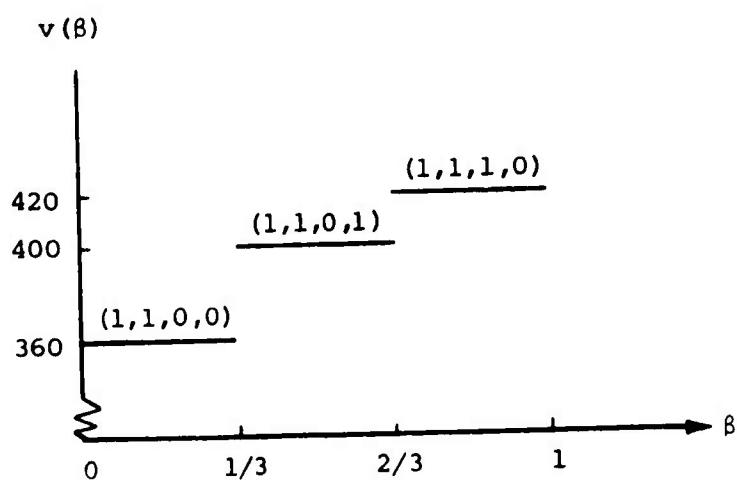
Figure 1.2.   Problem (1.1)--v as a Function of Parametric Right-Sides

13
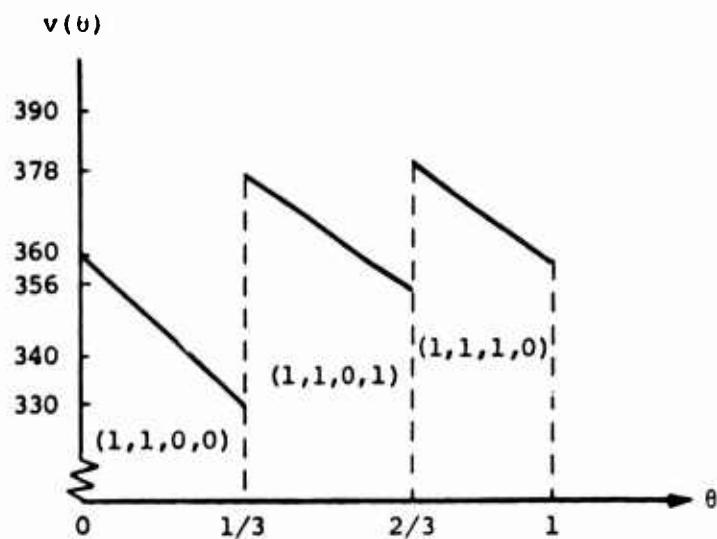
Figure 1.3.   Problem (1.1)--v as a Function of
Parametric Costs and Right-Sides

1.4 Conventional Means for Deriving Sensitivity Information

It is instructive to look briefly at various conventional methods
for deriving sensitivity information and to reflect upon these methods
with regard to their usefulness in the context of mixed integer pro-
gramming.   Four broad (not necessarily disjoint) approaches for deriving
information from a mathematical programming model are:

   (A)   Repetition of the optimization with different input data

   (B)   Quantitative analysis of solution output

   (C)   Post-optimal analysis

   (D)   Sequential run design.

The first method, repetition, involves making several runs with
perturbed data.   Sensitivity information is then gleaned from a compari-
son of the alternate solutions.   Repetition is common in simulations.
At times it may be desired only to compare alternate solutions with
regard to variables external to the model.   Application of parametric

14

programming techniques[*] could make this brute force method more attractive.

The second method is available in many optimization codes. In particular, the optimal dual variables in an LP problem describe marginal costs or benefits of implementing a nonoptimal policy or of modifying the constraints. In the LP case this information is free. In nonlinear (continuous-variable) problems that are solved via an LP-based algorithm, the dual variables (from the LP subproblem) can sometimes be transformed to represent meaningful sensitivities. Attempts to derive meaningful duals for MIP problems, however, have not been fruitful. It appears that an MIP analysis based solely on analysis of the solution output could produce only very limited sensitivity information. More will be said on this topic in later chapters.

Post-optimal analysis is distinguished from the second method in that it usually involves algorithmic modification, particularly of the final stage of the solution process. The usual style of post-optimal analysis is the derivation of sensitivities via (partial) reoptimization. Again using LP as an example, effects on the optimal value of slightly varying certain data, of fixing variables, and tightening or relaxing constraints are determined by reoptimization. This repetitive type of analysis is distinguished from method (A) in that it exploits the optimal solution and auxiliary output in order to hasten successive optimizations.

A post-optimal analysis employs techniques to efficiently reoptimize when slight perturbations are made to the original problem. The

---

[*] See Nauss [22].

15

efficiency of such analysis depends, of course, on the nature and magnitude of the perturbations to be studied; these may be such that the analysis procedure is degraded to the point of being nothing more than a repetition of runs (i.e., total reoptimizations). Unfortunately, the relation between efficiency of reoptimization and the "magnitude" of problem perturbations can usually be assessed only after considerable experience with similar problems has been gained.

Run design refers to the judicious sequencing of problems in order that the desired information is gained by executing an efficiently small number of runs. Successive problems are normally determined in sequential fashion and depend on the current status and type of information desired. The sequencing of problems should, of course, exploit reoptimization advantages by utilizing techniques of post-optimal analysis and parametric programming whenever applicable. Run design -- better known as experimental run design -- is popular in the analysis of problems of statistical nature; concepts of run design in the context of discrete optimization will be more clearly established and understood as the theory in Part II is unfolded (a good portion of the methodology presented in Part II can be regarded as run design).

These methods may be regarded as conventional when applied to continuous-variable problems, but in the treatment of discrete optimization problems there are presently no techniques (other than repetition) that can be called conventional. The state-of-the-art is discussed in paragraph 1.6.

1.5 Connections Between Continuity Theory, Sensitivity Analysis, and Parametric Programming

Continuity theory, the topic of Part I of this thesis, is the

16

formal basis for sensitivity analysis. It serves to justify many of the techniques examined in Part II. Naturally, the extensiveness and procedural details of a sensitivity analysis will depend on the inherent continuity properties of the model, provided that these properties are known or can be assessed prior to conducting the entire analysis.

Later on we will identify several broad classes of models that possess favorable continuity properties over all "interesting" data values. For these models it is likely a straightforward task to approximate or determine bounds on the fluctuation in the optimal value v with respect to changes in the data. In other words, models that are known or are determined to have nice continuity properties permit a different and usually less complicated analysis approach than do models that are known or are suspected to exhibit discontinuous behavior. Consequently, the approach that is taken to analyzing sensitivity will depend on an assessment of the continuity properties of the mode'.

Not only will the choice of approach depend on continuity properties of the model, but the approach itself may rely extensively on continuity considerations. In particular, issue (3) of paragraph 1.2 involves discovering data points[*] that are points of discontinuity.[†] Conditions for continuity/discontinuity derived in Part I will prove to be instrumental in constructing means for deriving points of discontinuity that permit a net improvement in optimal value.

The first issue stated in paragraph 1.2 is addressed completely by

_____

[*] A data point is a set of values for all data.

[†] In Figures 1.2 and 1.3, for instance, the points of discontinuity are of primary interest because they correspond to minimal data perturbations that can yield discrete improvements in optimal value.

17

the theory of continuity. If a discontinuity is recognized, the critical data and critical constraints can be identified by observing which data/constraints violate the conditions for continuity. Identification of critical data and constraints is particularly motivated by a need to reformulate the model.

Failing to establish continuity in a model, the analyst will want to identify points of discontinuity for the purposes of either altering the data to improve the objective value or of modifying the model toward greater realism. The task of identifying discontinuities or verifying that there are none in the domain of interest could be accomplished by systematically varying the problem data over the region of interest and solving the resulting collection of closely related problems.

Such a collection of related problems is referred to as a parametric integer program and the development/application of computational methods for efficiently solving a family of integer problems is referred to as parametric integer programming. Parametric programming in discrete optimization is a new field; the study by Nauss [22] of parametric integer linear programming is apparently the first substantial contribution to this field.

Thus, parametric programming can be viewed here as a technique to which the analyst can resort for analyzing continuity and sensitivity. However, it is a purpose of this treatise to present approaches to analyzing continuity and sensitivity that are far more efficient than a brute force repetition of problem solving, whether or not such repetition utilizes techniques of parametric programming. This is not to imply that efficient analysis techniques will be completely divorced from parametric programming; on the contrary, whenever an analysis

18

problem involves solving a sequence of related MIP problems, advantage should be taken of the applicable parametric programming techniques.

Typically parametric programming addresses parametric cost rows of the form $c + \gamma p$ and/or parametric right sides of the form $b + \beta r$ where γ and β are scalar parameters. In this context there are two categories of parametric programming considerations:

(A) Solve the finite number of similar MIP problems that correspond to specified values of the parameter.

(B) Explore the solutions that result when the parameter is allowed to vary continuously within the interval $[0,1]$.

A third consideration requires analysis of nonparametric data:

(C) Discover (nonparametric) data perturbations that permit net improved optimal solution values.

Parametric integer programming is limited primarily to category (A); the difficulty with (B) and (C) type problems is that a sequence of problems to be solved cannot be specified a priori. Techniques for handling these latter problems are particular topics to be addressed by the study of sensitivity theory and analysis.

It is not uncommon that parametric problems having type (B) or (C) objectives are recast into the type (A) format due to unavailability of appropriate analysis procedures. For instance, a manager may actually want to find (nonparametric) perturbed data values that yield substantial net improvements in optimal value, but for lack of an available (efficient) approach to the problem he invokes the brute force plan of solving the problem at several arbitrarily selected points.

The contact between sensitivity analysis and parametric programming can be made more explicit once the theory of sensitivity and analysis

19

procedures have been presented. Further comments on this topic are made in the chapters of Part II. Likewise, greater appreciation for the significance of continuity theory and its role in sensitivity analysis will be gained in Part II.

1.6 The State-of-the-Art

Theory and procedures for sensitivity analysis in (continuous) linear programming were developed over 20 years ago. Among the first treatments of this subject were Manne [20] and Gass and Saaty [10,11]. Comprehensive treatments of LP sensitivity analysis are given in Gass [9], Simmonard [28], and Dinkelbach [6], and basic LP sensitivity properties are offered in almost every LP text.

In the earliest attempts to derive integer solutions it was natural to exploit LP theory for any possible extensions and modifications that might lead to acceptable solution procedures. In many economic and production planning models the discrete variables are allowed to take on quite large values (e.g., 100); rounding an LP solution $x_i = 99.87$ to 100 likely is a good heuristic in this case, and in general, modifying an LP solution sometimes can provide a good suboptimal solution to the integer problem. When this is the case, LP sensitivity analysis can be put to use to estimate sensitivity in the integer problem. Along these lines is the work of Gomory [14], first published in 1958, and apparently the first serious treatment on the derivation and analysis of integer solutions. Similar modest beginnings in the analysis of integer programs that are based on (continuous) LP solutions include Frank [8], Jensen [19], and Aronofsky [2].

The limitations of any procedure based on LP solution modification

20

should be obvious, especially when applied to decision models with an abundance of 0-1 variables. It can be a trying task to massage an LP solution into one that is integer feasible, much less one that is near-optimal. In general, an attempt to derive an integer solution by modifying the LP solution must be assessed a failure whenever the numerical difference between the LP solution value and the manufactured integer solution value exceeds an acceptable tolerance of error. This being more often the case than not in discrete optimization problems, we must concede to the fact that LP-based sensitivity analysis provides little or no assistance in the analysis of discrete optimization problems.

The diff.culty in analyzing sensitivity in MIP problems is due primarily, of course, to the discontinuous behavior of v with respect to problem data, a property that rarely arises in continuous-variable LP problems. The size of the "jumps" in v and the proximity of the points of discontinuity are virtually impossible to predict. The analyst must resort to computational methods for discovering points of discontinuity and estimating the size of the discrete jumps.

Evidently, no procedures are presently available for sensitivity analysis of MIP problems, with the exception of the linear MIP problem with parametrized objective function. Since changes in objective function data do not influence the feasible region, v has characteristics much like those of the continuous-variable LP counterpart (piecewise-linear and concave) and can be studied via procedures derived for LP problems such as are described by Gass and Saaty [10,11]. A theoretical approach to estimating v for parametrized objective function data in the context of linear MIP is detailed by Noltemeier [23].

To the knowledge of this author, the first serious treatise on

21

sensitivity analysis for pure and mixed integer programming problems available in the literature was Noltemeier [23]. Except for the objective function parametrization mentioned above, however, the material is oriented toward abstract and theoretical (but elementary) properties that do not contribute to the more practical point-of-view taken here. No computational aspects are mentioned. Some of the results in [23] can be strengthened by imposing a boundedness assumption on the feasible variable space (which in practice is not really a restrictive assumption). A substantial portion of the theory concerns the "characteristic cone," which has no significance in the bounded case since it degenerates to a point (the origin). In any case, the pertinent theory presented in [23] is subsumed in the theory that we present in this treatise.

Sensitivity in integer programs is approached in Bowman [4] from a group theoretic point-of-view. The practicality of this approach has not been established, however.

A different kind of approach that can be considered within the realm of sensitivity analysis for integer programs is taken by Piper and Zoltners [24,25]. They describe an adjunct to implicit enumeration schemes that generates the K best ε-optimal solutions. Such a collection can be useful in considering variations in the data of the original problem.

Roodman [26,27] describes an adjunct to enumerative LP-based MIP algorithms that provides limited (single component) ranging analysis on particular parameters. Jensen [19] illustrates by way of a particular example the difficulties encountered when attempting a ranging analysis of an integer problem.

Another approach to obtaining MIP sensitivity information that appears to have been cast aside is the computation (reimputation) of useful dual prices. The first paper to address this topic was Gomory and Baumol [15] wherein a method was proposed for the distribution of dual values associated with cutting planes to the original constraints of the problem. The method is intimately tied to the method of integer forms of Gomory [14]. Unfortunately, the reimputed duals have severe shortcomings: their numerical values generally depend on the choice of cutting planes and a complete history of the added constraints is required for the recomputation; they display a sizable duality gap if interpreted as a dual solution; etc. A detailed critique of the Gomory and Baumol method is given by Weingartner [29]. Weingartner also presents an alternative approach in the context of capital budgeting problems that does not require record keeping of the cutting planes; the alternate duals are obtained from LP solutions and possess a greater degree of uniqueness than do the Gomory-Baumol duals, but they have a few other interpretive deficiencies. Alcaly and Klevorick [1] also present some methods by which the duals can be improved to be more meaningful. Hespel [16] summarizes the accomplishment in this area and offers still another improvement, but yet severe discrepancies in the reimputed dual prices persist and meaningful interpretation is open to question. All of these treatments are based on Gomory's method of integer forms for solving integer programming problems. Additional comments on this topic are offered in Appendix 10A.

Except for recent work by Williams [31], the study of continuity in MIP problems has been limited to the reimputation and interpretation of dual prices. Williams establishes some basic continuity results and

23

recommends penalty functions as a remedy for handling constraints that are likely to allow unrealistic discontinuities. No procedures or implementation guidelines are given, however.

A considerable amount of literature is available dealing with continuity of linear and nonlinear continuous-variable programs. Foremost in this area are the papers by Evans and Gould [7]; Meyer [21]; Hogan [18]; Dantzig, Folkman and Shapiro [5]. These studies exploit the use of point-to-set mapping theory established primarily by Zangwill [32,33] and Berge [3]. Williams [30] relates a form of directional continuity to boundedness of the primal and dual solution sets (and to the equivalent regularity conditions); Hoffman and Karp [17] establish LP continuity with respect to all parameters in terms of the same conditions.

Application of point-to-set mapping theory to mathematical programming is given by Hogan [18]. Conditions establishing continuity of extremal value functions and properties of maps determined by inequalities are included.

1.7 The Plan of the Paper

The treatment of sensitivity analysis in mixed integer programming is presented in two parts. Continuity theory is addressed in Part I, which serves as a foundation upon which to build analysis procedures. Sensitivity theory and analysis techniques are presented in Part II. The main emphasis is on sensitivity to right-side data. We will elaborate only briefly on the general content. A more detailed synopsis is given in the introduction to each part.

24

The bulk of the material applies to nonlinear MIP problems in general, but care is taken not to overlook special properties of linear problems. Evidently, much of the theory of continuity that applies to linear MIP problems can be extended to convex programs; nevertheless, in order to encapsulate the linear theory a special chapter in Part I is dev⸱⸱ continuity in linear MIP problems. Likewise, the analysis considerations of Part II apply in general to nonlinear problems and any particular advantages that occur in the linear case are pointed out.

The theory of continuity is based on point-to-set mapping theory [3,18,32,33]. Thus, the set of feasible variable values is viewed as a mapping of the data point (i.e., the data defining the model) into the feasible region. Continuity of the optimal value and optimal solution set as functions of the data is expressed in terms of set-interiority conditions and continuity properties of the constraint and objective functions.

Sensitivity analysis theory and procedures presented in Part II emphasize three classes of data: objective function data, parametric right-side data, and nonparametric right-side data. Sensitivity analysis with respect to nonparametric data permits arbitrary perturbation of the data. The trite consideration given to constraint function data is justified by the observation that the major concerns with this data can be addressed in terms of the right-side data.

BIBLIOGRAPHY - CHAPTER 1

1. Alcaly, R.E., and A.K. Klevorick, "A Note on the Dual Prices of
    Integer Programs," Econometrica, Vol. 34, No. 1 (January 1966).

2. Aronofsky, J.S. (ed.), Progress in Operations Research, Vol. III,
    John Wiley, New York, 1969.

3. Berge, C., Topological Spaces, Macmillan, New York, 1963.

4. Bowman, V.J., "The Structure of Integer Programs under the Hermitian
    Normal Form," Operations Research, Vol. 22, No. 5 (September-
    October 1974).

5. Dantzig, G.B., J. Folkman, and N. Shapiro, "On the Continuity of
    the Minimum Set of a Continuous Function," J. Math. Anal. and Appl.,
    Vol. 17 (1967).

6. Dinkelbach, W., Sensitivitätsanalyse und parametrische Programm-
    mierung, Springer-Verlag, New York, 1969.

7. Evans, J.P., and F.J. Gould, "Stability in Nonlinear Programming,"
    Operations Research, Vol. 18, No. 1 (January-February 1970).

8. Frank, C.R., "Parametric Programming in Integers," Operations-
    Research-Verfahren, Bd. III (1967).

9. Gass, S.I., Linear Programming, McGraw-Hill, New York, 1958.

10. _____, and T.L. Saaty, "A Computational Algorithm for the Para-
    metric Objective Function," Naval Research Logistics Quarterly,
    Vol. 2 (1955).

11. _____, and _____, "Parametric Objective Function - Part II:
    Generalization," Operations Research, Vol. 3, No. 4 (1955).

12. Geoffrion, A.M., "A Guided Tour of Recent Practical Advances in
    Integer Linear Programming," Working Paper No. 220, Western Manage-
    ment Science Institute, UCLA, revised May 1975.

13. _____, and R.E. Marsten, "Integer Programming Algorithms: A
    Framework and State-of-the-Art Survey," Management Science, Vol.
    18, No. 9 (May 1972).

14. Gomory, R.E., "Outline of an Algorithm for Integer Solutions to
    Linear Programs," Bulletin of the American Mathematical Society,
    64, September 1958.

15. _____, and W. Baumol, "Integer Programming and Pricing," Econo-
    metrica, Vol. 28, No. 3 (July 1960).

16. Hespel, P.P., "Reimputing the Dual Variables in Integer Programming," Discussion Paper No. 6707, Center for Operations Research and Econometrics, Universite Catholique de Louvain, Heverlee, Belgium, May 1967.

17. Hoffman, A.J., and R.M. Karp, "On Nonterminating Stochastic Games," Management Science, Vol. 12, No. 5 (January 1966), Appendix.

18. Hogan, W.W., Jr., "Point-to-Set Maps in Mathematical Programming," SIAM Review, Vol. 15, No. 3 (July 1973).

19. Jensen, R.E., "Sensitivity Analysis and Integer Linear Programming," The Accounting Review (July 1968).

20. Manne, A.S., "Notes on Parametric Linear Programming," Report No. P-468, The Rand Corporation, Santa Monica, California, 1953.

21. Meyer, R., "The Validity of a Family of Optimization Methods," SIAM J. Control, Vol. 8, No. 1 (February 1970).

22. Nauss, R.M., "Parametric Integer Programming," Working Paper No. 226, Western Management Science Institute, UCLA, January 1975.

23. Noltemeier, H., "Sensitivitätsanalyse bei diskreten linearen Optimierungsproblemen," Lecture Notes in Operations Research and Mathematical Systems, #30, M. Beckman and H.P. Kunzi (eds.), Springer-Verlag, New York, 1970.

24. Piper, C.J., and A.A. Zoltners, "Implicit Enumeration Based Algorithms for Postoptimizing Zero-One Programs," Management Sciences Research Report No. 313, Graduate School of Industrial Administration, Carnegie-Mellon University, March 1973.

25. _____, and _____, "A Pragmatic Approach to 0-1 Decision Making," Management Sciences Research Report No. 316, Graduate School of Industrial Administration, Carnegie-Mellon University, April 1973.

26. Roodman, G.M., "Postoptimality Analysis in Zero-One Programming by Implicit Enumeration," Naval Research Logistics Quarterly, Vol. 19 (1972).

27. _____, "Postoptimality Analysis in Integer Programming by Implicit Enumeration: The Mixed Integer Case," Working Paper, Amos Tuck School of Business Administration, Dartmouth College, October 1973.

28. Simmonard, M., Linear Programming, Prentice-Hall, Englewood Cliffs, New Jersey, 1966.

29. Weingartner, H.M., Mathematical Programming and the Analysis of Capital Budgeting Problems, Prentice-Hall, Englewood Cliffs, New Jersey, 1963.

30. Williams, A.C., "Marginal Values in Linear Programming," <u>SIAM J. App. Math.</u>, Vol. 11, No. 1 (March 1963).

31. _____, "Some Modeling Principles for M.I.P.'s," paper presented at the VIII International Symposium on Mathematical Programming, Stanford University, August 1973.

32. Zangwill, W.I., "Convergence Conditions for Nonlinear Programming Algorithms," <u>Management Science</u>, Vol. 16, No. 1 (September 1969).

33. _____, <u>Nonlinear Programming: A Unified Approach</u>, Prentice-Hall, Englewood Cliffs, New Jersey, 1969.

PART I

CONTINUITY THEORY

Chapters 2-7

- Introduction to Part I
- Fundamental Continuity Properties of Nonlinear MIP Problems
- Interiority Conditions of Continuity in Nonlinear MIP Problems
- Continuity Properties of Convex MIP Problems
- Continuity Properties of Linear MIP Problems
- Some Notes on Model Formulation

29

CHAPTER 2

INTRODUCTION TO PART I

A mathematical program with integer restrictions on some of the variables is much more likely to exhibit discontinuous behavior in optimal value with respect to problem data than the counterpart problem where none of the variables are restricted to be integer. When a small perturbation in the data results in an incommensurate change in the optimal value, the analyst will likely want to alter the data and/or the model. If the discontinuity is realistic, he may want to slightly alter the data in order to discretely improve the objective; if unrealistic, he may need to improve the model artfully so that it adequately represents the real problem. When a program is determined to be continuous at the given data point the analyst is then in a position to determine further sensitivity properties of the optimal value.

The analyst using MIP models and codes will want to extend his interest beyond the derivation of an optimal solution and value: typically, he will be interested first in knowing which data changes, if any, can yield discrete changes in the optimal value. Investigation of this question will be referred to as continuity analysis, the topic to which Part I is devoted.

2.1 Content and Summary

Conditions are developed for the general bounded mixed integer programming (MIP) problem that assure continuity at a point. Point-to-set mapping theory plays a central role: with reasonably weak assumptions on the objective function, a discontinuity in optimal value is possible only when the feasible region is discontinuous as a point-to-

30

set map. When this map is "closed" the optimal value is a semicontinuous function of the data so that a small change in data cannot create an incommensurate improvement in optimal value; this property always holds for bounded linear MIP problems. Continuity in the linear case is assured by a single post-optimal set-interiority condition.

The fundamental conditions for continuity in MIP problems are developed in Chapter 3. The treatment by Hogan [16] of point-to-set maps in mathematical programming is the prime source of theory upon which the first results of this chapter are based. Also included in this chapter is a result concerning directional continuity that has an important implication in continuity analysis; essentially, this result states that the optimal value v is continuous with respect to all data if it is continuous in a certain direction with respect to the right-side data.

The condition that a point lie in the interior or relative interior of its domain is referred to as interiority. Two domains are of interest: the set of points in variable space which are feasible in the MIP problem at the specified data point (i.e., the feasible region), and the set of data points for which the corresponding feasible region is not empty. The latter set is referred to as the "data set." If continuity holds at a data point α in the continuous-variable problem that results when the discrete variables are held fixed at optimal values, and if α is in the interior of the data set, then v is continuous with respect to the data at the point α. A similar statement holds when data set interiority is replaced by the stronger condition of feasible region interiority, which will be referred to as "strong feasibility." Interiority conditions for continuity are the subject of Chapter 4.

31

The results apply in general to nonlinear MIP problems.

Specializations to the convex and linear cases are made in Chapters 5 and 6, respectively. The conditions for continuity simplify especially in the linear case bacause of the favorable continuity properties of LP problems. A summary of LP continuity properties is offered in Chapter 6.

Currently, there is a great deal of interest in linear MIP and comparatively little in nonlinear MIP; the reasons are understandable. But this does not mean that the usefulness of sensitivity theory can be construed with a proportionate bias toward the linear case. The primary support for this claim is the fact that linear MIP problems are frequently the result of a linearization process applied to a nonlinear program. Granted, the analyst is concerned with sensitivity properties of the derived linear MIP, but he is also interested in the properties of the original program. In this case the analyst requires only a linear MIP code but he desires more general sensitivity theory.

In further support of studying nonlinear theory, the evolution of specialized (e.g., convex) nonlinear MIP codes places greater demand on a general theory of continuity and associated computational sensitivity analysis procedures. More importantly, the prior existence of such theory could prove to be of great assistance to the modeler as he attempts to construct a realistic formulation of the problem.

Much of the content of Chapters 3, 4, 5 and 6 is directed toward the development of conditions which assure that v is continuous, i.e., sufficient conditions. Conditions that establish a discontinuity, on the other hand, are not as accessible. Violation of the sufficient conditions for continuity, however, is strong evidence of a discontinuity.

Of course, assurance of a discontinuity can be gained by computational measures if necessary. In some cases, conditions are expressed that are both necessary and sufficient.

Once confronted with a discontinuity, the analyst is obliged to explain the bizarre behavior. Is the model an inaccurate representation of the real world? What data and/or which constraints are responsible for causing the discontinuity? How can the model be made more realistic? Quite probably some of the apprehension in the use of MIP codes and models is due to the fact that so little is known about MIP continuity properties and their implications in modeling and computation.

Chapter 6 exploits the implications of continuity theory in the formulation of MIP models. In the formulation of a linear capital investment problem, for example, a budgetary limitation is imposed as a constraint which expresses that the total capital invested cannot exceed a prescribed number, say $\bar{b}$. Thus, any investment plan is precluded which exceeds the budget $\bar{b}$ by an arbitrarily small amount. Williams [22] points out that such a model does not reflect the real world fact that a budget can always be stretched a little if just a small additional amount of capital is needed to provide a much better investment plan. Evidently, when integer restrictions are placed on some of the variables, greater care must be given to the formulation of an MIP problem in order to achieve an acceptable representation of the real world.

To recapitulate, Part I (Chapters 3 through 7) addresses the following primary considerations of continuity analysis:

(A) Can a small change in data yield a substantial change in the objective value?

(B) What data and which constraints are responsible for a discontinuity?

(C) How may the model be improved?

2.2 Terminology and Definitions

We address the following generic MIP problem:

$$
(I) \begin{cases} \text{minimize } f(\alpha;x,y) \\ \text{subject to: } (x,y) \in B(\alpha) \\ \qquad\qquad\qquad x \in X \\ \qquad\qquad\qquad y \in Y \end{cases} \tag{2.1}
$$

where: α denotes the vector of data ($\alpha \in E^r$)

   x   denotes the continuous variables

   y   denotes the discrete variables

   X   is a compact[*] set in E^p

   Y   is a finite[*] set in E^q

and $f : E^r \times X \times Y \to E$.

The <u>optimal value</u> of problem (I) at α is denoted by $v(\alpha)$. Thus,[†]

$$
v(\alpha) \equiv \underset{x \in X, y \in Y}{\text{minimum }} f(\alpha;x,y) \,|\, (x,y) \in B(\alpha) \ . \tag{2.2}
$$

The set of <u>ε-optimal solutions</u> at α is denoted by

$$
M_\varepsilon(\alpha) \equiv \{ (x,y) \in F(\alpha) \,|\, f(\alpha;x,y) \leq v(\alpha) + \varepsilon \} \tag{2.3}
$$

---

[*] These are not severely restrictive assumptions in practice since practical variable bounds are usually obvious.

[†] "Minimum" in place of "infimum" is permissible because of boundedness.

where the feasible region, F(α), is given by

$$F(\alpha) \equiv B(\alpha) \cap (X \times Y) .$$

The set A of all values α which admit a feasible solution to problem (I) is given by

$$A \equiv \{\alpha \in E^r | F(\alpha) \neq \emptyset\} . \tag{2.4}$$

When the y-variables are held fixed at $\overline{y} \in Y$, the corresponding subproblem is given as

$$(I_{\overline{y}}) \begin{cases} \text{minimize } f(\alpha;x,\overline{y}) \\ \quad x \in X \\ \text{subject to: } (x,\overline{y}) \in B(\alpha) . \end{cases}$$

Accordingly, we define

$$v_{\overline{y}}(\alpha) \equiv \text{optimal value of problem } (I_{\overline{y}}) \text{ at } \alpha$$

$$F_{\overline{y}}(\alpha) \equiv \{x \in X | (x,\overline{y}) \in F(\alpha)\} \tag{2.5}$$

$$A_{\overline{y}} \equiv \{\alpha \in A | F_{\overline{y}}(\alpha) \neq \emptyset\} .$$

Notice that $v_{\overline{y}}(\alpha) \geq v(\alpha)$, $F_{\overline{y}}(\alpha) \subseteq F(\alpha)$, and $A_{\overline{y}} \subseteq A$.

The expression int(A) denotes the _interior_ of A relative to E^r. If $\alpha \in \text{int}(A)$ then there is a neighborhood N(α), an open set in E^r containing α, such that $N(\alpha) \subseteq A$. The relation $\alpha \in \text{int}(A_y)$ means that there is a neighborhood about α in which y remains feasible. The expression bd(A) denotes the boundary of A, i.e., the closure of A, denoted cl(A), less int(A).

The set A will be said to be of full dimension if it has a non-empty interior relative to E^r. One would expect that in most instances A is full-dimensional since it is dictated only by feasibility. But suppose the analyst must impose certain "specification" constraints on

35

his selection of values for the data α. Such constraints involving only α (and not x or y) naturally do not appear in problem (I) and therefore do not influence the set A. For example, suppose that a constraint characterizing $B(\alpha)$ is

$$\alpha_1 x_1 + \alpha_2 x_2 \leq 1$$

where the analyst selected values for α_1 and α_2 such that $\alpha_1 + \alpha_2 = 1$. If it is imperative that these data sum to one, then we should be cognizant of that fact when studying the continuity of v with respect to α. Specifically, we need only to address continuity relative to the subset of A for which $\alpha_1 + \alpha_2 = 1$. Only this subset is "of interest" to the analyst.

We shall let $\tilde{A}$ denote the set of values for α that are of interest to the analyst. To distinguish between all data and data "of interest," the elements of $\tilde{A}$ will be referred to as parameters. Typically, the analyst is interested only in values of α in proximity to a nominal point $\bar{\alpha}$ and which satisfy the specifications for selecting the parameters α. Additionally, $\tilde{A}$ may be restricted by a dependence of α on auxiliary parameters β: $\alpha = \alpha(\beta)$ where it is assumed, of course, that $\alpha(\cdot)$ is continuous. For example, consider the function (i.e., either an objective or constraint function)

$$\alpha_1 x_1 + \alpha_2 x_2$$

where $\alpha_1 = \beta$ and $\alpha_2 = \beta^2$. If α_1 and α_2 are to be treated as individual parameters (alternatively, α_1 and α_2 can be replaced by the single parameter β) then $\tilde{A}$ is restricted by the condition that $\alpha_2 = (\alpha_1)^2$. On the other hand, the relation between α_1 and α_2 might be fuzzy in the

sense that the analyst is willing to entertain values for α_1 and α_2 given by

$$\alpha_1 = (1 \pm \epsilon_1)\beta \quad \text{and} \quad \alpha_2 = (1 \pm \epsilon_2)\beta^2 .$$

In this case, α_1 and α_2 must be viewed as individual parameters with no restrictions imposed upon them.

What this discussion leads to is the fact that we want to address the continuity of v relative to a subset of A (which is possibly of smaller dimension than A). To this end, we define

$$\hat{A} \equiv A \cap \tilde{A}$$

$$\hat{A}_y \equiv A_y \cap \tilde{A} . \tag{2.6}$$

These sets are illustrated in two dimensions in Figure 2.1.
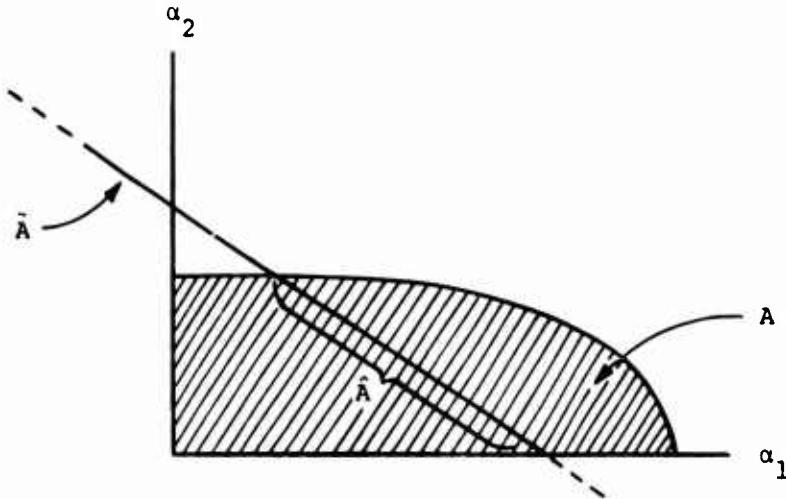
Figure 2.1.   The Sets A, $\tilde{A}$, and $\hat{A}$

Definition 1: $v(\cdot)$ is <u>ϵ-continuous relative to $\hat{A}$</u> at $\hat{\alpha} \in \hat{A}$ if for any $\gamma > 0$ there exists a neighborhood $N_\delta(\hat{\alpha}) \subseteq E^r$ such that

$$\alpha \in \hat{A} \cap N_\delta(\hat{\alpha}) \implies |v(\hat{\alpha}) - v(\alpha)| < \gamma + \epsilon .$$

$v(\cdot)$ is <u>continuous</u> (i.e., 0-continuous) at $\hat{\alpha}$ if it is continuous relative

37

to E^r at $\hat{\alpha}$; $v(\cdot)$ is continuous on a set S if it is continuous relative to S at each $\alpha \in$ S.

Lower and upper semi-continuity, abbreviated l.s.c. and u.s.c., respectively, are weaker than continuity. In this paper they are applied only to functions.

<u>Definition 2</u>: A function $g(t)$ is l.s.c. at t if for $\gamma > 0$ there is a neighborhood $N_\delta(t)$ such that

$$t' \in N_\delta(t) \Rightarrow g(t) < g(t') + \gamma .$$

$g(\cdot)$ is u.s.c. if $-g(\cdot)$ is l.s.c.

We will see that $v(\cdot)$ is l.s.c. under very mild conditions.

Another kind of continuity that will be useful is referred to as "directional" continuity. It is well-known that a function can be discontinuous at a point even though it is continuous along all lines through the point. For example, the function given as

$$g(x,y) \equiv \begin{cases} 0 & \text{if } y = 0 \\ x^2/y & \text{if } y \neq 0 \end{cases}$$

is continuous along all lines through the origin, $y = ax$: for $x \neq 0$, $g(x,ax) = \dfrac{x}{a} \to 0$ as $x \to 0$, and $g(0,0) = 0$. On the other hand, along the curves $y = x^2$, we have $g(x,x^2) = 1$ for $x \neq 0$, and $g(0,0) = 0$; hence, g is not continuous at $(0,0)$.

We shall say that this function is directionally continuous although it is not continuous.

<u>Definition 3</u>: A function $g(\cdot)$ is <u>continuous in direction ξ</u> at t if for any $\gamma > 0$ there exists a $\delta > 0$ such that

$$0 \leq \theta \leq \delta \Rightarrow \left| g(t + \theta\xi) - g(t) \right| \leq \gamma$$

38

where t and ξ have the same dimension.

Of course, continuity implies directional continuity. A case where the converse holds is with the function

$$v(b) = \min\, f(x,y)\,\big|\,G(x,y) \le b$$

where f and G are continuous in x for fixed y. It will be seen in the next section that v(b) is not necessarily continuous at b, but if it is continuous in a certain direction, then it is continuous.

We will need to address continuity of functions of the form $g(\alpha;x,y)$ on the domain $A \times X \times Y$. But each element y of Y is discrete so continuity on, for instance, $\alpha \times X \times y$ makes no sense. What we need is continuity of $g(\cdot;\cdot,y)$ as a function of α and x on $\alpha \times X$ for y fixed. We shall use the "at" symbol, @, as follows.

Definition 4: "g continuous on $\alpha \times X$ @ y" will be taken to mean that $g(\cdot;\cdot,y)$ is continuous on $\alpha \times X$ for y held fixed; g is continuous on $\alpha \times X$ @ Y if g is continuous on $\alpha \times X$ @ y for each $y \in Y$.

Actually, we will address continuity on $\alpha \times F_y(\alpha)$ rather than the larger set $\alpha \times X$. Of course, y must be such that $F_y(\alpha) \ne \emptyset$.

Definition 5: $y \in Y$ is _feasible at α_ if $F_y(\alpha) \ne \emptyset$. Thus, y is feasible at α if there is an $x \in X$ such that $(x,y) \in B(\alpha)$.

Definition 6: The set of y's feasible at α is denoted by $Y(\alpha)$. Thus

$$Y(\alpha) \equiv \{y \in Y \mid F_y(\alpha) \ne \emptyset\}.$$

Analogously, the subset of X for which (x,y) is feasible for some $y \in Y$ is given by

$$X(\alpha) \equiv \{x \in X \mid F_y(\alpha) \ne \emptyset \text{ for some } y \in Y\}.$$

The sets $X(\alpha)$ and $Y(\alpha)$ satisfy $X(\alpha) \times Y(\alpha) = F(\alpha)$. Notice also that $X(\alpha) = \bigcup_Y F_y(\alpha)$.

Finally, we will need the concept of continuity as it applies to point-to-set maps. The entities $F(\cdot)$, $B(\cdot)$, and $M_\epsilon(\cdot)$ which we have already defined can be regarded as point-to-set maps. For example, F maps a point α into the set of points $F(\alpha)$; thus $F:A \to X \times Y$. Let T denote a general map, $T:A \to Z$ where $A \subseteq E^r$ and $Z \subseteq E^n$.

Definition 7: The map T is open (relative to A) at $\alpha \in A$ if $\{\alpha^k\} \subset A$, $\alpha^k \to \alpha$, and $z \in T(\alpha)$ imply the existence of an integer m and a sequence $\{z^k\} \subset Z$ such that $z^k \in T(\alpha^k)$ for $k \geq m$ and $z^k \to z$.

Definition 8: The map T is closed (relative to A) at $\alpha \in A$ if $\{\alpha^k\} \subset A$, $\alpha^k \to \alpha$, $z^k \in T(\alpha^k)$, and $z^k \to z$ imply that $z \in T(\alpha)$.

Definition 9: T is continuous (relative to A) at $\alpha \in A$ if it is both open and closed (relative to A) at α.

In the literature[*] the terms lower and upper semicontinuous used in the context of point-to-set maps have meanings similar to open and closed, respectively. In fact, for the spaces which we shall deal with here, these similar terms are equivalent. We choose, therefore, to use the terms lower and upper semicontinuous (l.s.c. and u.s.c.) only as they pertain to functions.

A geometric representation of open and closed maps is perhaps easier to achieve by visualizing the sequence of sets $T(\alpha^k)$ where $\alpha^k \to \alpha$. To wit, it is obvious that T is open at α if $T(\alpha) \subseteq \lim T(\alpha^k)$ and T is closed at α if $T(\alpha) \supseteq \lim T(\alpha^k)$. It follows that T is

_____

[*] For instance: Berge [4]; Dantzig, et al. [6]; Debreu [8]; Evans and Gould [9]; Hogan [16]; Meyer [17]; Zangwill [23,24].

40

continuous at α if $T(\alpha) = \lim T(\alpha^k)$. Conditions which are necessary as well as sufficient are stated without proof as follows:

T is open at α if and only if $T(\alpha) \subseteq$ closure $[\lim T(\alpha^k)]$.

T is closed at α if and only if closure $[T(\alpha)] \supseteq \lim T(\alpha^k)$.

T is continuous at α if and only if closure $[T(\alpha) =$

closure $[\lim T(\alpha^k)]$.

Figure 2.2(a) depicts a closed map that is not open and Figure 2.2(b) an open map that is not closed.

Capital letters are used to denote maps, sets, and vector-valued functions. Scalar-valued functions and points are denoted by lower case letters. Vector inequalities apply to each component.
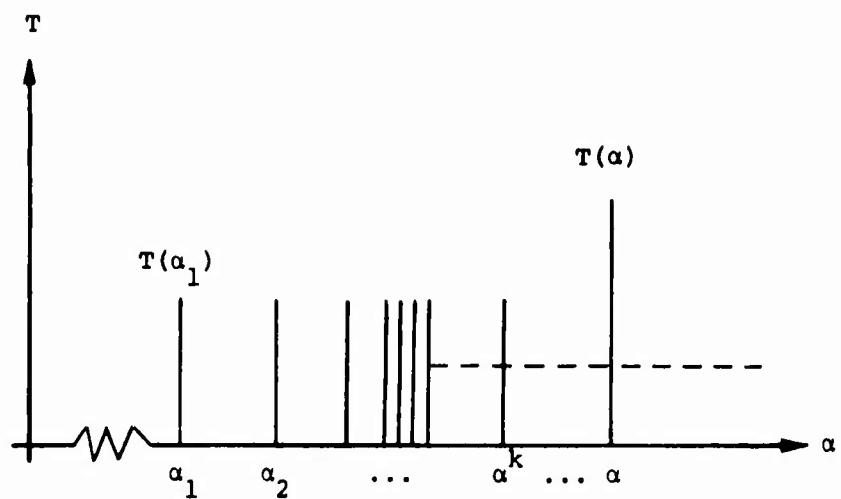
Figure 2.2(a). A Closed Map That is Not Open
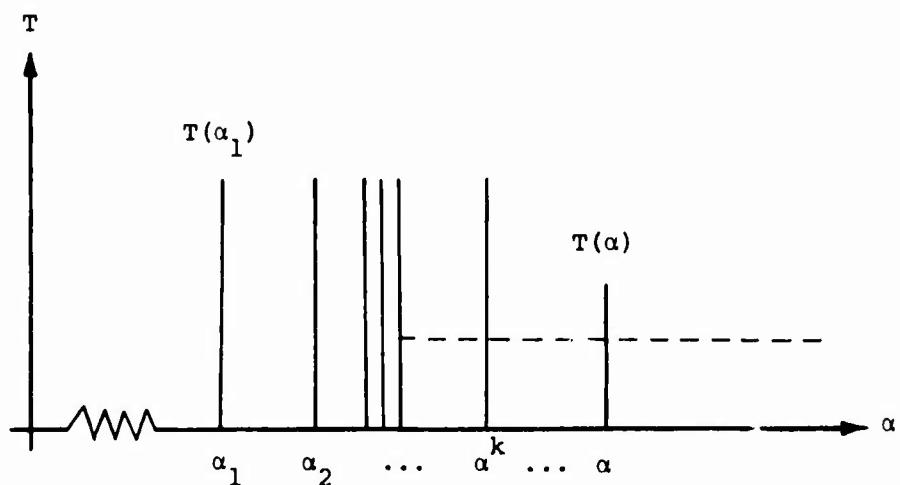
Figure 2.2(b). An Open Map That is Not Closed

CHAPTER 3

FUNDAMENTAL CONTINUITY PROPERTIES OF NONLINEAR PROBLEMS

Conditions are derived which assure that $v(\cdot)$ and/or $M(\cdot)$ are con-
tinuous with respect to the data α. Certain continuity properties of
$v(\cdot)$ and $M(\cdot)$ are assured by similar properties of the objective and
constraint functions. Stronger continuity results, however, depend on
an "interiority condition" that the data vector α be in the interior of
a certain set of parameters. A more restrictive condition, referred to
as "strong feasibility" by Williams [22], is also studied.

This chapter addresses the general bounded MIP problem defined as
problem (2.1) in Chapter 2. Application to the linear case is deferred
to Chapter 6. Thus, we will address the optimal value function given
by

$$v(\alpha) \equiv \underset{x \in X, y \in Y}{\text{minimum}} \ f(\alpha;x,y) \mid (x,y) \in B(\alpha) \qquad (3.1)$$

where X is compact and Y is a finite set. Other point-to-set maps of
interest are the feasible region, $F(\alpha) \equiv B(\alpha) \cap (X \times Y)$, and the set of
solutions, $M_\varepsilon(\alpha)$, that are ε-optimal at α ($\varepsilon \geq 0$).

Not all data may be of concern in continuity analysis; only certain
data, e.g., some of the right-side data and/or some of the cost data,
are usually of interest. The "interesting" data will be referred to as
parameters.

The more detailed results of this chapter pertain to the charac-
terization of $B(\alpha)$ as a collection of functional constraints on the
variables. However, several statements can be made about the continu-
ity of $v(\cdot)$ and $M(\cdot)$ with regard to the completely general domain,
$B(\alpha)$. Before giving a characterization of $B(\alpha)$ (hence, $F(\alpha)$) we offer

the following has. ? conditions for continuity.

3.1 Basic Results From the Theory of Point-to-Set Maps

Theorem 3.1 is taken directly from Hogan [16] and expressed in the context of problem (3.1). The statements of this theorem are fundamental in the development to follow.

<u>Theorem 3.1</u>: (a) If $F(\cdot)$ is closed at $\bar{\alpha} \in A$, if there exists a neighborhood N of $\bar{\alpha}$ such that the closure of $\underset{\alpha \in N}{\cup} F(\alpha)$ is compact, and if f is l.s.c. on $\bar{\alpha} \times X(\bar{\alpha})$ @ $Y(\bar{\alpha})$, then $v(\cdot)$ is l.s.c. at $\bar{\alpha}$.

(b) If $F(\cdot)$ is open at $\bar{\alpha} \in A$ and f is u.s.c. on $\bar{\alpha} \times X(\bar{\alpha})$ @ $Y(\bar{\alpha})$, then $v(\cdot)$ is u.s.c. at $\bar{\alpha}$.

(c) If $F(\cdot)$ is continuous at $\bar{\alpha} \in A$ and f is continuous on $\bar{\alpha} \times X(\bar{\alpha})$ @ $Y(\bar{\alpha})$, then $M_\epsilon(\cdot)$ is closed at $\bar{\alpha}$.

Not much is lost in the strength of these statements if the domain $\bar{\alpha} \times X(\bar{\alpha})$ @ $Y(\bar{\alpha})$ is replaced by $\bar{\alpha} \times X$ @ Y. The former is generally not available for assessment of the hypothesized conditions; besides, in most practical mathematical programming problems continuity of f holds over the entire domain $\bar{\alpha} \times X$ @ Y. The next four theorems obtain easily from Theorem 3.1.

<u>Theorem 3.2</u>: Given that $\alpha \in A$, if

  (1) f is l.s.c. on $\alpha \times X(\alpha)$ @ $Y(\alpha)$, and

  (2) $F_y(\cdot)$ is closed at α for each $y \in Y(\alpha)$ (equivalently, if
       $F(\cdot)$ is closed at α),

then $v(\cdot)$ is l.s.c. at α.

<u>Proof</u>: Since Y is finite, $F(\cdot) = \underset{Y}{\cup} (F_y(\cdot) \times y)$ is a closed map if each $F_y(\cdot)$ is closed. Furthermore, $F(\cdot)$ is uniformly bounded since $X \times Y$ is

44

compact. Thus, f and F satisfy the conditions of Theorem 3.1(a) and the result obtains. ||

If problem (2.1) were stated as a maximization problem, then we must replace "l.s.c." in Theorem 3.2 by "u.s.c.," and vice versa in the next theorem.

With regard only to continuity of v, lower semicontinuity is, in one sense, all that really matters to the (naive) analyst, for it means that small perturbations in α cannot cause a discrete improvement in the infimal value v. (We will see shortly that conditions for $F(\cdot)$ to be closed are very mild so that by Theorem 3.2, v is l.s.c. at α also under mild conditions.) On the other hand, it would be shortsighted to conclude that a continuity analysis can be limited to an investigation of lower semicontinuity. The analyst interested in varying parameters must be concerned with any discontinuity, whether it causes a discrete jump up or a jump down in the infimal value. Moreover, if v is l.s.c. at α but not continuous, there is some question as to whether the model realistically reflects the true problem.

Theorem 3.3: Given that $\alpha \in A$, if

   (1) f is u.s.c. on $\alpha \times X(\alpha)$ @ $Y(\alpha)$, and

   (2) $F_y(\cdot)$ is open at α for each $y \in Y(\alpha)$ (equivalently, if $F(\cdot)$ is open at α),

then $v(\cdot)$ is u.s.c. at α.

Proof: $F(\cdot)$ is open since the union of an arbitrary family of open maps is open. The result follows since $F(\cdot)$ and f satisfy the conditions of Theorem 3.1(b). ||

Combining these two theorems, we have the following continuity

45

result.

Theorem 3.4: Given $\alpha \in A$, if

    (1) f is continuous on $\alpha \times X(\alpha) @ Y(\alpha)$, and

    (2) $F_y(\cdot)$ is continuous at α for each $y \in Y(\alpha)$ (equivalently,

        if $F(\cdot)$ is continuous at α),

then $v(\cdot)$ is continuous at α.

The conditions of Theorem 3.4 also imply that the map $M(\cdot)$, which

identifies the set of optimal solutions, is closed at α. This follows

from Theorem 3.1(c).

Theorem 3.5: Given $\alpha \in A$ and $\varepsilon \geq 0$, (1) and (2) of Theorem 3.4 imply

$M_\varepsilon(\cdot)$ is closed at α.

An interesting special class of mixed integer programs is one in

which all parameters appear only in the objective function. In this

case, $F(\cdot)$ is continuous since it is constant with respect to parameter

changes. By Theorem 3.4, v is continuous on A provided only that the

objective function is continuous on $A \times X @ Y$.

To be more general, we observe that v is continuous with respect

to all components of α which appear only in the objective function if

f is continuous with respect to α. Going one step further, we conjecture

that if $F(\cdot)$ is continuous with respect to all other components of α,

then v is continuous at $\overline{\alpha}$ (i.e., with respect to all components). That

this conjecture is indeed valid is the content of the following theorem.

Theorem 3.6: Let $\alpha = (\alpha^1; \alpha^2)$ where α^1 appears only in the objective

function, f, which is assumed to be continuous on $\overline{\alpha} \times X(\alpha) @ Y(\overline{\alpha})$.

Then v is continuous at $\overline{\alpha}$ if v is continuous with respect to α^2 at $\overline{\alpha}$.

Proof: We are given that v is continuous with respect to α^1 at $\bar{\alpha}$ and with respect to α^2 at $\bar{\alpha}$. That is, $v(\alpha^1;\bar{\alpha}^2)$ is continuous as a function of α^1 at $\bar{\alpha}^1$, and $v(\bar{\alpha}^1;\alpha^2)$ is continuous as a function of α^2 at $\bar{\alpha}^2$. First, we note that f is uniformly continuous with respect to x @ y since $X \times Y$ is compact. That is, for $\epsilon > 0$ there exists a $\delta_1 > 0$ such that

$$\left| \alpha^1 - \bar{\alpha}^1 \right| \leq \delta_1 \Rightarrow \left| f(\alpha^1;\bar{\alpha}^2;x,y) - f(\bar{\alpha};x,y) \right| \leq \frac{\epsilon}{2} \tag{3.2}$$

for any $x \in X(\bar{\alpha})$ and $y \in Y(\bar{\alpha})$. Secondly, by continuity of $v(\alpha^1;\bar{\alpha}^2)$, there exists a $\delta_2 > 0$ such that

$$\left| \alpha^1 - \bar{\alpha}^1 \right| \leq \delta_2 \Rightarrow \left| v(\alpha^1;\bar{\alpha}^2) - v(\bar{\alpha}) \right| \leq \frac{\epsilon}{2} . \tag{3.3}$$

Thirdly, by continuity of $v(\bar{\alpha}^1;\alpha^2)$, there exists a δ_3 such that

$$\left\| \alpha^2 - \bar{\alpha}^2 \right\| \leq \delta_3 \Rightarrow \left| v(\bar{\alpha}^1;\alpha^2) - v(\bar{\alpha}) \right| \leq \frac{\epsilon}{2} . \tag{3.4}$$

Define $\delta = \min(\delta_1,\delta_2,\delta_3)$ and let α be a generic point satisfying $\left| \alpha - \bar{\alpha} \right| \leq \delta$ so that conditions (3.2), (3.3), and (3.4) all hold at α.

Next, observe that if (x^*,y^*) is any optimal solution point at $(\bar{\alpha}^1;\alpha^2)$, then (x^*,y^*) is feasible at $(\alpha^1;\alpha^2) = \alpha$ so that

$$\begin{aligned} v(\alpha) &\leq f(\alpha;x^*,y^*) \\ &\leq v(\bar{\alpha}^1;\alpha^2) + \frac{\epsilon}{2} \qquad \text{[by (3.2)]} \\ &\leq v(\bar{\alpha}) + \epsilon \qquad \text{[by (3.3)]} . \end{aligned} \tag{3.5}$$

Similarly, if $(\hat{x},\hat{y})$ is optimal at α, then it is feasible at $(\bar{\alpha}^1,\alpha^2)$ so that

$$\begin{aligned} v(\bar{\alpha}^1;\alpha^2) &\leq f(\bar{\alpha}^1;\alpha^2;\hat{x},\hat{y}) \\ &\leq f(\alpha;\hat{x},\hat{y}) + \frac{\epsilon}{2} \qquad \text{[by (3.2)]} \\ &= v(\alpha) + \frac{\epsilon}{2} . \end{aligned} \tag{3.6}$$

Since $v(\bar{\alpha}) - \epsilon/2 \leq v(\bar{\alpha}^1;\alpha^2)$ by (3.4), we have

$$v(\overline{\alpha}) - \epsilon \leq v(\alpha) \ . \qquad\qquad (3.7)$$

Finally, combining (3.5) and (3.7),

$$v(\overline{\alpha}) - \epsilon \leq v(\alpha) \leq v(\overline{\alpha}) + \epsilon$$

which establishes continuity of v with respect to α at $\overline{\alpha}$. $\qquad$ ||

Example 1: The economic lot size model with N periods and equal demand D in each period can be expressed as

$$\text{minimize} \sum_{k=1}^{N} (c_k x_k + s_k y_k)$$

$$\text{subject to:} \sum_{i=1}^{k} x_i \geq kD, \qquad k = 1,2,\ldots,N$$

$$0 \leq x_k \leq y_k(\text{ND}), \quad k = 1,2,\ldots,N$$

$$y_k = 0 \text{ or } 1, \qquad k = 1,2,\ldots,N \ .$$

The parameters include D, and the c_k and s_k, $k = 1,2,\ldots,N$. By substituting Dz_k for x_k, we obtain an equivalent formulation:

$$\text{minimize} \sum_{k=1}^{N} [(c_k D)z_k + s_k y_k]$$

$$\text{subject to:} \sum_{i=1}^{k} z_i \geq k, \qquad k = 1,2,\ldots,N$$

$$0 \leq z_k \leq Ny_k, \qquad k = 1,2,\ldots,N$$

$$y_k = 0 \text{ or } 1, \qquad k = 1,2,\ldots,N \ .$$

We observe that in this formulation all parameters appear only in the objective function and we conclude that v is continuous with respect to D, c_k, and s_k, $k = 1,2,\ldots,N$ (where $D \geq 0$, $c_k \geq 0$, $s_k \geq 0$).

Example 2: Let $a \geq 0$ and c be scalars and consider the hypothetical problem:

$$\max y - cx$$

$$\text{subject to:} \quad (1 + a)\, x \geq y$$

$$0 \leq x \leq y$$

$$y = 0 \text{ or } 1 .$$

The infimal value v is continuous with respect to c (i.e., as a function of c alone) since c appears only in the objective function. The feasible region is given by

$$F(a) = \{(0,0)\} \cup [1/(1 + a),\, 1] \times \{1\}$$

which clearly varies continuously with respect to a (a > 0). Hence, by Theorem 3.4, v is continuous with respect to a (i.e., as a function of a alone) at any value a > 0. By Theorem 3.6, v is (jointly) continuous with respect to c and a (for any values c and a > 0).

It will be useful to recognize that there are two kinds of discontinuity (relative to E^r) that can arise: those caused by discontinuity of F(·) relative to A, and those caused by bordering infeasibility (α ε bd(A)) where we define v(α) = ∞ for α ∉ A (for minimization problems). A point of discontinuity α will be said to be of Category I if v(·) is discontinuous relative to A at α. A point of discontinuity α will be said to be of Category II if α ε bd(A).

A Category I discontinuity is characterized by a finite jump in the optimal value v. In the capital investment example problem (1.1) of Chapter 1 we saw that a budget of b = (80,70) yields a value v = 360, but a budget b_1 less than 80 by an arbitrarily small amount yields a jump down in optimal value to v = 300. The point b = (80,70) is clearly an interior point in the feasible parameter set A; therefore it is Category I.

A Category II discontinuity is characterized by an infinite jump since $|v| = \infty$ for $\alpha \not\in A$. In the problem

$$\text{maximize } x$$
$$x \geq 0$$

$$\text{subject to: } x \leq \alpha$$

we have $A = [0,\infty)$, and

$$v(\alpha) = \begin{cases} \alpha, & \alpha \geq 0 \\ -\infty, & \alpha < 0 . \end{cases}$$

The point $\alpha = 0$ is a Category II discontinuity. Notice that v is continuous on A so that $\alpha = 0$ is not of Category I. In fact, in any bounded LP problem, a point of discontinuity must be of Category II since the optimal value is continuous on int(A) (see Chapter 6).

The following observations are all quite obvious and are stated without proof:

(1) Every point of discontinuity is either Category I, II, or both.

(2) A point of discontinuity α is only Category I if and only if $\alpha \in \text{int}(A)$.

(3) If f is finite on $A \times X \times Y$, then every point on bd(A) and only these points are Category II points of discontinuity.

(4) If v is finite and ε-continuous relative to E^r at α, then $\alpha \in \text{int}(A)$.

Evidently, to investigate the continuity of v at a point $\alpha \in A$ we must consider two things: the continuity of v relative to A at α, and the fact that α is or is not in the interior of A (relative to E^r). If A has an interior and α is contained in this interior, then continuity and relative continuity at α are equivalent. If $\alpha \in \text{bd}(A)$, which

50

includes the case where int(A) = ∅, then we are concerned only with the particular perturbations of α that are feasible and we seek to determine if α is also a Category I point of discontinuity.

Recall that the parameters α are completely general (see paragraph 2.2 regarding A and Â) and may be viewed as continuous functions of auxiliary parameters. For instance, cost coefficients in a capacity expansion model may depend on discount factors that in turn depend on interest rates in each period. If α is a continuous (vector) function of other parameters β, then v is continuous with respect to β if v is continuous with respect to α.

In general, we are interested in continuity of v relative to Â or to the sets $\hat{A}_y$. Of course, continuity of v relative to Â is implied if v is continuous relative to A. Typically, the set A is easier to characterize than Â, so it makes sense to first investigate continuity relative to A; then if v proves to be discontinuous relative to A at $\bar{\alpha}$ we will want to determine if discontinuity obtains only for perturbational changes (directions) from $\bar{\alpha}$ which are prohibited by the requirement that α be restricted to Â.

One more observation worth mentioning regards uniform continuity and boundedness of v. If the set Â of interesting and permissible values for α is a compact subset of A which has an interior relative to E^r, then v(·) continuous on Â implies that v(·) is uniformly continuous and bounded on Â. In this case, continuity on Â is "smooth" in the sense that there can be no cusps or points with unbounded slopes.

3.2 Continuity Conditions on the Constraint Functions

The set B(α) is described here in terms of a collection of inequality

51

and equality constraints, viz.:

$$B(\alpha) = \{(x,y) \mid G(\alpha;x,y) \leq 0, \ H(\alpha;x,y) = 0\} \qquad (3.8)$$

where $G:A \times X \times Y \rightarrow E^m$ and $H:A \times X \times Y \rightarrow E^n$. Thus, G has m component functions $g_1,\ldots,g_m$ and H has n component functions $h_1,\ldots,h_n$. The distinction between these two types of constraints is made because the conditions on G and H which assure that $B(\cdot)$ is closed are different. To wit, we state the following lemmas.

Lemma 3.7: $P(\alpha) \equiv \{(x,y) \mid x\varepsilon X, \ y\varepsilon Y, \ G(\alpha;x,y) \leq 0\}$ is a closed map at α if each component function of G is l.s.c. on $\alpha \times X(\alpha) \ @ \ Y(\alpha)$.

Proof: Given $\alpha^k \rightarrow \alpha$, $(x^k,y^k) \rightarrow (x,y)$, $G(\alpha^k;x^k,y^k) \leq 0$. Let $\varepsilon > 0$ be an arbitrary m-vector. By lower semicontinuity of G, there is a K_ε such that $k \geq K_\varepsilon$ implies

$$G(\alpha;x,y) \leq G(\alpha^k;x^k,y^k) + \varepsilon \leq \varepsilon \ .$$

Since ε is arbitrary, $G(\alpha;x,y) \leq 0$ and $P(\cdot)$ is closed at α. $\qquad ||$

Notice that since Y is finite, $y^k \rightarrow y$ requires that there be a K_0 such that $y^k = y$ for all $k \geq K_0$. Notice also that G l.s.c. on $\alpha \times X \ @ \ Y(\alpha)$ is equivalent to G l.s.c. on each subset, $\alpha \times X \ @ \ y$, $y \ \varepsilon \ Y(\alpha)$.

Lemma 3.8: $Q(\alpha) = \{(x,y) \mid x\varepsilon X, \ y\varepsilon Y, \ H(\alpha;x,y) = 0\}$ is a closed map at α if each component of H is continuous on $\alpha \times X(\alpha) \ @ \ Y(\alpha)$ (i.e., continuous on $\alpha \times X \ @ \ y$ for each y feasible at α).

Proof: H and -H are l.s.c. on $\alpha \times X \ @ \ Y$, so the maps $Q^-(\alpha) \equiv \{(x,y) \mid x\varepsilon X, \ y\varepsilon Y, \ H(\alpha;x,y) \leq 0\}$ and $Q^+(\alpha) \equiv \{(x,y) \mid x\varepsilon X, \ y\varepsilon Y, \ H(\alpha;x,y) \geq 0\}$ are closed at α by Lemma 3.7. Hence, $Q = Q^- \cap Q^+$ is closed at α. $\qquad ||$

With $B(\cdot)$ given by (3.8), problem (2.1) appears as the following:

$$\text{Minimize } f(\alpha;x,y)$$
$$x\epsilon X, y\epsilon Y$$

$$\text{s.t.} \quad G(\alpha;x,y) \leq 0 \tag{3.9}$$

$$H(\alpha;x,y) = 0 \ .$$

The feasible region is given by

$$F(\alpha) = P(\alpha) \cap Q(\alpha) = \{(x,y) \,|\, x\epsilon X, \ y\epsilon Y, \ G(\alpha;x,y) \leq 0, \ H(\alpha;x,y) = 0\} \ .$$

It is important to recognize that problem (3.9) permits the generalities that components of α can be interdependent and that any component of α can appear in several or all of the objective and constraint functions. Some later developments will demonstrate how improved results can be obtained by sacrificing generality in representation of the data α.

Theorem 3.9: If G is l.s.c. on $\alpha \times X(\alpha) \ @ \ Y(\alpha)$ and H is continuous on $\alpha \times X(\alpha) \ @ \ Y(\alpha)$, then $F(\cdot)$ is closed at α.

Proof: $F(\cdot)$ is the finite intersection of closed maps (by Lemmas 3.7 and 3.8). $||$

Theorem 3.10: If f is l.s.c., G l.s.c., and H continuous on $\alpha \times X(\alpha) \ @ \ Y(\alpha)$ then $v(\cdot)$ is l.s.c. at α.

Proof: (If $\alpha \not\in A$, $v(\cdot) \equiv +\infty$.) $F(\cdot)$ is closed at α by Theorem 3.9, and the result follows from Theorem 3.2.

Theorem 3.10 states that v cannot "jump down" in value due to a "small" change in α. That is, for any $\epsilon > 0$ there is a neighborhood about α such that the infimal value at any point in this neighborhood is not less than $v(\alpha) - \epsilon$.

This result is almost intuitive when we consider that $F(\cdot)$ closed

at α means that the feasible region cannot discretely increase in size when α is only slightly perturbed. In particular, this asserts that a discrete value of variable y which is infeasible at α cannot be feasible at any point arbitrarily close to α. This observation is formalized as follows.

Theorem 3.11: If $F(\cdot)$ is closed at $\overline{\alpha}$, then there exists a neighborhood $N(\overline{\alpha})$ about $\overline{\alpha}$ such that $Y(\alpha) \subseteq Y(\overline{\alpha})$ for all $\alpha \in N(\overline{\alpha})$.

Proof: Notice that the set $F_y(\overline{\alpha})$ is closed for any y. Thus, $F_y(\cdot)$ closed at $\overline{\alpha}$ means that $\lim F_y(\alpha^k) \subseteq F_y(\overline{\alpha})$ for any sequence $\alpha^k \to \overline{\alpha}$. If y is infeasible at $\overline{\alpha}$, then $F_y(\overline{\alpha}) = \emptyset$ which implies that $F_y(\alpha^k) = \emptyset$ for infinitely many k. Thus, there exists a neighborhood about $\overline{\alpha}$ in which y is infeasible. The result follows since Y is a finite set. $||$

In order to apply Theorem 3.4 to establish continuity of v, we would like to determine conditions under which $F(\cdot)$ is open. Unfortunately, these conditions are more severe and more difficult to obtain than those for closedness of $F(\cdot)$. Furthermore, $F(\cdot)$ open is not necessary. For cases where $F(\cdot)$ is not open, we are interested in knowing what conditions in addition to interiority are sufficient for $v(\cdot)$ to be continuous. This is the topic of Chapter 4.

3.3 Directional Continuity

Oftentimes it is reasonable to consider only "directional" changes in the data; this can make the analysis of continuity a much simpler task. The term directional is used rather loosely to designate data changes that are restricted in some way. For instance: increases in objective function data, all other data held fixed; increases in right-side data, all other data fixed; or changes in right-side data

satisfying b + θr where θ is a scalar.

Consider the special case α = (c,a,b) where the objective function f depends only on c, G depends only on a, b denotes right-side data, and c, a, and b are independent. This special case of problem (3.9) is expressed as

$$
\begin{array}{l}
\text{minimize } f(c;x,y) \\
x\epsilon X, y\epsilon Y
\end{array}
\qquad (3.10)
$$

$$
\text{subject to: } G(a;x,y) \leq b .
$$

Equality constraints could be included in this problem provided they are assumed independent of a, b, and c (i.e., $H(x,y) = 0$).

One continuity property of problem (3.10) is immediate: as a function of c alone, v is continuous for all values of c. As a special case, the optimal value function

$$
\hat{v}_c(\theta) \equiv \text{minimum } f(c+\theta p;x,y) \,|\, (x,y) \,\epsilon\, F
$$

is continuous on E (the real line).

For ease of exposition, define

$$
\overline{v}(b) \equiv v(\overline{a},b,\overline{c}), \qquad \overline{a} \text{ and } \overline{c} \text{ fixed}
$$

$$
\overline{v}(a,b) \equiv v(a,b,\overline{c}), \qquad \overline{c} \text{ fixed}
$$

$$
\overline{v}(c) \equiv v(\overline{a},\overline{b},c), \qquad \overline{a} \text{ and } \overline{b} \text{ fixed} .
$$

As before, $\hat{A} = A \cap \tilde{A}$ where $A \equiv \{(a,b) \,|\, (3.10) \text{ is feasible}\}$ and $\tilde{A}$ denotes any specification constraints on a and b.

Given that f is continuous on $(\overline{a},\overline{b},\overline{c}) \times X \text{ @ } Y$, it follows by Theorem 3.6 that $v(a,b,c)$ is continuous at $(\overline{a},\overline{b},\overline{c})$ if $\overline{v}(a,b)$ is continuous at $(\overline{a},\overline{b})$. We will show that if there exists a direction $r < 0$ along which $\overline{v}(b)$ is continuous (at $\overline{b}$), then $\overline{v}(a,b)$ is continuous at $(\overline{a},\overline{b})$ provided that G is continuous on $\overline{a} \times X \text{ @ } Y$.

55

This proposition is fairly intuitive: if $\bar{b}$ is perturbed in any way that increases the feasible region, then $\bar{v}(\bar{b})$ cannot increase in value and by Theorem 3.10 it cannot jump down in value. Continuity in a direction $r < 0$ therefore assures that $\bar{v}(b)$ is continuous at $\bar{b}$. Let "a" denote a (small) perturbation of $\bar{a}$. If (x,y) is feasible at $(\bar{a},\bar{b})$, then by continuity of G there is a "b" close to $\bar{b}$ such that (x,y) is feasible at (a,b). Continuity of $\bar{v}(b)$ therefore implies continuity of $\bar{v}(a,b)$. A detailed proof is given below.

Define the function of a single variable:

$$\hat{v}(\theta) \equiv \begin{cases} \text{minimum } f(\bar{c},x,y) \\ x \epsilon X, y \epsilon Y \\ \text{subject to: } G(\bar{a};x,y) \leq \bar{b} + \theta \cdot \bar{r} \end{cases} \qquad (3.12)$$

where θ is a scalar and $\bar{r}$ is an arbitrarily chosen negative vector: $\bar{r} < 0$.

<u>Theorem 3.12</u>: If f is continuous on $\bar{c} \times X @ Y$ and G is continuous on $\bar{a} \times X @ Y$, then $v(a,b,c)$ is continuous at $(\bar{a},\bar{b},\bar{c})$ if and only if $\hat{v}(\cdot)$ is right-continuous at 0.

<u>Proof</u>: (Necessity is obvious.) $\hat{v}$ is l.s.c. at 0 by Theorem 3.10; that is, given $\epsilon > 0$ there exists a $\gamma > 0$ such that $\hat{v}(0) \leq \hat{v}(\theta) + \epsilon$ whenever $|\theta| \leq \gamma$. Notice that for $\theta \leq 0$, $\hat{v}(\theta) \leq \hat{v}(0)$. Hence, $-\gamma < \theta \leq 0 \Rightarrow \hat{v}(0) - \epsilon \leq \hat{v}(\theta) \leq \hat{v}(0)$ so that $\hat{v}$ is left-continuous at 0. With the hypothesis, we have that $\hat{v}$ is continuous at 0. Thus, for $\epsilon > 0$ there exists a $\delta > 0$ such that $|\theta| \leq \delta$ implies

$$|\hat{v}(\theta) - \hat{v}(0)| \leq \epsilon/2 . \qquad (3.13)$$

We will show next that $v(b)$ is continuous at $\bar{b}$. By (3.13) and monotonicity of $\hat{v}$, we may write

56

$$0 \leq \dot{v}(\delta) - \dot{v}(-\delta) = \overline{v}(\overline{b} + \delta \cdot \overline{r}) - \overline{v}(\overline{b} - \delta \cdot \overline{r}) \leq \epsilon \ . \qquad (3.14)$$

(Remember that $\overline{r} < 0$.) The hypercube $[\overline{b} + \delta \cdot \overline{r}, \ \overline{b} - \delta \cdot \overline{r}]$ defines a neighborhood about $\overline{b}$ such for any $\breve{b}$ in this neighborhood, i.e., for

$$\overline{b} + \delta \cdot \overline{r} \leq \breve{b} \leq \overline{b} - \delta \cdot \overline{r}$$

we have by monotonicity of $\overline{v}(b)$ that

$$\overline{v}(\overline{b} + \delta \cdot \overline{r}) \geq \overline{v}(\breve{b}) \geq \overline{v}(\overline{b} - \delta \cdot \overline{r}) \ .$$

Similarly,

$$\overline{v}(\overline{b} + \delta \cdot \overline{r}) \geq \overline{v}(\overline{b}) \geq \overline{v}(\overline{b} - \delta \cdot \overline{r})$$

so that by (3.14),

$$\left| \overline{v}(\breve{b}) - \overline{v}(\overline{b}) \right| \leq \epsilon \ . \qquad (3.15)$$

This establishes continuity of $\overline{v}(b)$ at $\overline{b}$. We next establish continuity of $\overline{v}(a,b)$ at $(\overline{a},\overline{b})$. Since $\overline{v}(a,b)$ is l.s.c. at $(\overline{a},\overline{b})$ by Theorem 3.10, we need only to establish that

$$\overline{v}(a,b) \leq \overline{v}(\overline{a},\overline{b}) + \epsilon \qquad (3.16)$$

for (a,b) sufficiently close to $(\overline{a},\overline{b})$. Suppose that (x^*,y^*) is an optimal solution at $(\overline{a},\overline{b}+\delta \cdot \overline{r})$. Then

$$G(\overline{a};x^*,y^*) \leq \overline{b} + \delta \cdot \overline{r} < \overline{b} \ . \qquad (3.17)$$

Define

$$N(\overline{a}) \equiv \{a \mid G(a;x^*,y^*) \leq G(\overline{a};x^*,y^*) - \frac{1}{2}\delta \cdot \overline{r}\} \qquad (3.18)$$

$$N(\overline{b}) \equiv [\overline{b} + \frac{1}{2}\delta \cdot r, \ \infty) \ . \qquad (3.19)$$

Clearly, $N(\overline{a}) \times N(\overline{b})$ is a neighborhood of $(\overline{a},\overline{b})$. Take $(\tilde{a},\breve{b}) \ \epsilon \ N(\overline{a}) \times N(\overline{b})$. Then (x^*,y^*) is feasible at $(\tilde{a},\breve{b})$ since

57

$$G(\tilde{a};x^*,y^*) \le G(\bar{a};x^*,y^*) - \frac{1}{2}\delta \cdot \bar{r} \qquad \text{[by (3.18)]}$$

$$\le \bar{b} + \frac{1}{2}\delta \cdot \bar{r} \qquad \text{[by (3.17)]}$$

$$\le \tilde{b} \qquad \text{[by (3.19)]} .$$

Thus, for any $(\tilde{a},\tilde{b}) \in N(\bar{a}) \times N(\bar{b})$,

$$\bar{v}(\tilde{a},\tilde{b}) \le \bar{v}(\overline{a,b} + \delta \cdot \bar{r}) \qquad \text{[by feasibility of } (x^*,y^*)\text{]}$$

$$\le \bar{v}(\bar{a},\bar{b}) + \varepsilon \qquad \text{[by (3.15)]} .$$

This establishes (3.16) so that $\bar{v}(a,b)$ is continuous at $(\bar{a},\bar{b})$. Also, $\bar{v}(c)$ is continuous at $\bar{c}$ by Theorem 3.4 since $F(\cdot)$ is independent of c. Finally, $v(a,b,c)$ is continuous at $(\bar{a},\bar{b},\bar{c})$ by Theorem 3.6. $||$

The conclusion in Theorem 3.12 may be stated as: $v(a,b,c)$ is continuous relative to $\hat{A}$ at (a,b,c) if $\hat{v}(\cdot)$ is right-continuous relative to $\hat{A}$ at 0. This results by simply replacing $N(\bar{a}) \times N(\bar{b})$ by $\hat{A} \cap [N(\bar{a}) \times N(\bar{b})]$ in the proof.

The practical significance of this result stems from the fact that it is much easier to observe or compute that v changes continuously with respect to decrementing b than it is to determine that v is continuous with respect to all parameters. It is sufficient in problem (3.10) to determine that v does or does not change (increase) by a "small" amount when each component of b is decremented by a "small" amount.

CHAPTER 4

INTERIORITY CONDITIONS FOR CONTINUITY IN NONLINEAR PROBLEMS

The condition that a point lie in the interior of its domain is referred to as interiority. Two kinds of set interiority will be studied: parameter or data set interiority (i.e., $\alpha \in \text{int}(A)$), and feasible region interiority (which will be defined as "strong feasibility"). This chapter is devoted to the mathematical expression of interiority conditions and derivation of additional conditions which are sufficient for continuity of $v(\cdot)$ and/or closedness of $M_\epsilon(\cdot)$.

It is especially difficult with mixed integer programs to determine "a priori," i.e., by analysis of the model itself (without the aid of any particular solutions), that $v(\cdot)$ is continuous. This would require observing that the problem possesses desirable properties for each feasible $y \in Y$, since it is not known prior to solution which value(s) of y is optimal. It is certainly less effort to study the continuity properties of a problem when y is fixed at some feasible value. Assessment of problem properties via inspection of the optimal or ϵ-optimal solutions is referred to as "post-optimal continuity analysis." Post-optimal analysis usually involves assessment of continuity properties only at an optimal solution point.

Once again we will be addressing problem (3.9). Thus,

$$v(\alpha) \equiv \underset{x \in X, y \in Y}{\text{minimum}} f(\alpha;x,y) \,|\, G(\alpha;x,y) \leq 0, \; H(\alpha;x,y) = 0 \,.$$

Some of the following results address continuity on the subset A_y, $y \in Y$, and are therefore most applicable as a post-optimal analysis where y is optimal (or ϵ-optimal). But is should be recognized that

59

these statements apply to a generic subset A_y and if the stated conditions can be assessed for all feasible $y \in Y$, then an a priori conclusion can be made. The sets $\hat{A}$ and $\hat{A}_y$ may be read in place of A and A_y, respectively, in the material to follow.

Before giving a counterpart to Theorem 3.10, we will need to introduce some additional interiority terms. Illustrations of the concepts defined herewith are intended to cultivate insight into geometric interpretations of subsequent results.

4.1 Set Interiority Concepts and Definitions

The _interior_ of a set is a relative term; it must be used with reference to a given space or set. The entity int(A) was defined earlier to be the interior of A relative to E^r, since A is defined as a set in E^r. Thus,

$$\text{int}(A) \equiv E^r - \text{cl}(E^r - A)$$

$$(4.1)$$

where
$$\text{cl}(\cdot) \equiv \text{closure } (\cdot)$$

($\hat{A}$, A_y, or $\hat{A}_y$ may be read in place of A). Another set operator is "ri," which denotes the _relative interior_. The relative interior of a set A is the set of all points $\alpha \in A$ such that for any $\epsilon > 0$ there is a δ-neighborhood, $\delta \leq \epsilon$, given by

$$N_\delta(\alpha) \equiv \{\alpha' \in E^r \mid \|\alpha - \alpha'\| < \delta\} \qquad (4.2)$$

which contains at least two points $\alpha^1 \in A$ and $\alpha^2 \in A$ for which $\|\alpha^1 - \alpha^2\| > \delta$. Thus,

$$\text{ri}(A) \equiv \{\alpha \in A \mid \text{ for any } \epsilon > 0 \text{ and some } \delta \leq \epsilon$$

$$(4.3)$$

$$\exists \; \alpha^1 \text{ and } \alpha^2 \in A \cap N_\delta(\alpha) \ni \|\alpha^1 - \alpha^2\| > \delta\} \; .$$

A subset, e.g., A_y, of A will be said to be of the _same dimension_

as A if there is an open set N in E^r such that $A \cap N \subseteq A_y$. A set in E^r is of __full dimension__ if it has the same dimension as E^r; thus, A is of full dimension if it has a nonempty interior (this definition was given earlier).

The terms "int," "cl," "ri," and "same dimension" defined in the context of convex sets are used by Rockafellar [19]. The sets A, $\hat{A}$, A_y, and $\hat{A}_y$, however, are not necessarily convex, even for linear MIP problems.

__Example:__ Consider a problem with two parameters (i.e., $r = 2$) and suppose that $A = [0,1] \times [0,1]$ and $A_y = [0,1] \times \{1/2\}$. These sets are depicted in Figure 4.1(a). We have

$$\text{int}(A) = (0,1) \times (0,1)$$

$$\text{int}(A_y) = \emptyset$$

$$\text{ri}(A) = (0,1) \times (0,1)$$

$$\text{ri}(A_y) = (0,1) \times \{1/2\} \ .$$

If A_y were given as $[0,1] \times [0,1/2]$ as shown in Figure 4.1(b), then we have

$$\text{int}(A_y) = (0,1) \times (0,1/2)$$

$$\text{ri}(A_y) = (0,1) \times (0,1/2) \ .$$

If the dimension of parameter space were given as $r = 3$, then $\text{int}(A) = \text{int}(A_y) = \emptyset$ but $\text{ri}(A)$ and $\text{ri}(A_y)$ remain as given above (in both situations cited).
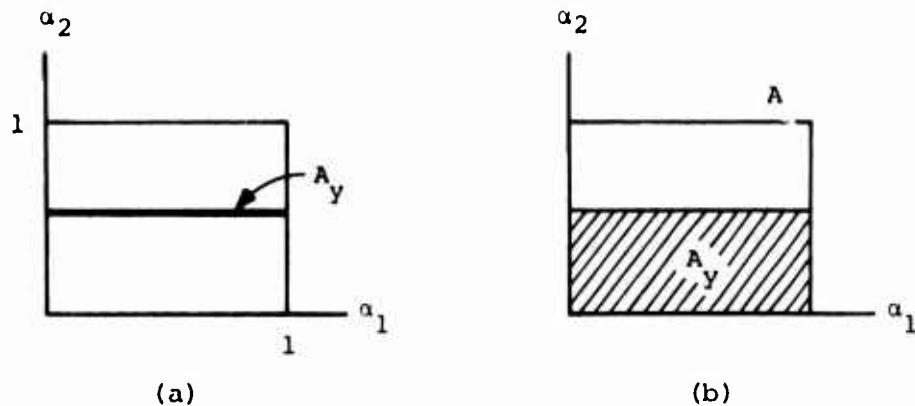
Figure 4.1. Example Sets A and A_y in E^2

The example demonstrates the obvious fact that $ri(A) = int(A)$ if the set A is of full dimension. It is also obvious (since $A_y \subseteq A$) that

$$int(A_y) \subseteq int(A)$$

and (4.4)

$$cl(A_y) \subseteq cl(A)$$

but it is not true in general that $ri(A_y) \subseteq ri(A)$. Take, for instance, $A = [0,1] \times [0,1]$ and $A_y = [0,1] \times \{0\}$. The relative interiors are, in fact, disjoint: $ri(A) = (0,1) \times (0,1)$ and $ri(A_y) = (0,1) \times \{0\}$.

The interiority conditions for continuity that we shall present shortly could be adequately expressed in terms of the interior and relative interior set operators; however, the results are somewhat stronger and the conditions more digestible when we utilize a concept that we shall describe as the interior of a subset relative to its containing set. The interior of A_y relative to A (or $\hat{A}_y$ relative to $\hat{A}$) is defined as

$$int(A_y | A) \equiv \{\alpha \varepsilon \, ri(A_y \mid \exists \, N(\alpha) \ni A \cap N(\alpha) \subseteq A_y\}$$ (4.5)

62

where $N(\alpha)$ is an arbitrary (open) neighborhood about α. It should be clear that

$$
'~ \iota(A_y|A) = \begin{cases} \text{int}(A_y), & \text{if } A \text{ and } A_y \text{ are of full dimension} \\ \text{ri}(A_y), & \text{if } A \text{ and } A_y \text{ are of the same dimension} \\ \emptyset, & \text{if } A_y \text{ is not of the same dimension as } A \text{ .} \end{cases} \qquad (4.6)
$$

Finally, the set that we shall employ is slightly larger than $\text{int}(A_y|A)$ in that it includes the common boundary points of A_y and A. For this reason, we shall call this set the pseudo-interior of A_y relative to A; it is defined (quite simply) as

$$
\text{pi}(A_y|A) \equiv \{\alpha \in A_y \mid \exists~ N(\alpha) \ni A \cap N(\alpha) \subseteq A_y\} \text{ .} \qquad (4.7)
$$

Example: Suppose $\hat{A}$ is a closed planar disc in E^3 and that $\hat{A}_y$ is half of the disc shown by the shaded region in Figure 4.2(a). Thus, $\text{int}(\hat{A}) = \text{int}(\hat{A}_y) = \emptyset$. The set $\text{int}(\hat{A}_y|\hat{A})$ is the open half disc and $\text{pi}(\hat{A}_y|\hat{A})$ is the open half disc plus the boundary portion of $\hat{A}$ that is common to $\hat{A}_y$. These sets are depicted in Figure 4.2(b).
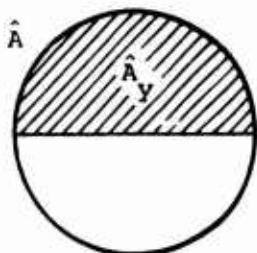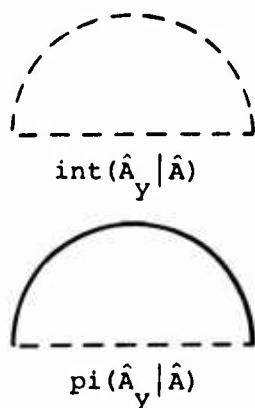
(a)   (b)

Figure 4.2. Example Interior and Pseudo-Interior Sets

4.2 Parameter Set Interiority

We are now able to give fairly tight sufficiency conditions for
continuity of v. Once again, the sets $\hat{A}$ and $\hat{A}_y$ may be read in place of
A and A_y, respectively, in all of the following statements.

Theorem 4.1: Let $\overline{v}$ be optimal at $\overline{\alpha}$ in problem (3.9). If

(a) $\overline{\alpha} \in pi(A_{\overline{y}}|A)$, and

(b) $v_{\overline{y}}$ is finite and u.s.c. relative to $A_{\overline{y}}$ at $\overline{\alpha}$

then v is u.s.c. relative to A at $\overline{\alpha}$.

Proof: (a) implies that there is a neighborhood $N_\delta(\overline{\alpha})$ such that y is
feasible at any $\alpha \in A \cap N_\delta(\overline{\alpha})$. By (b), for $\varepsilon > 0$ there exists a δ^1,
$0 < \delta^1 \leq \delta$, such that $\alpha \in A \cap N_{\delta^1}(\overline{\alpha})$ implies that

$$v_{\overline{y}}(\overline{\alpha}) > v_{\overline{y}}(\alpha) - \varepsilon \ ,$$

or

$$v_{\overline{y}}(\alpha) < v(\overline{\alpha}) + \varepsilon$$

since $v_{\overline{y}}(\overline{\alpha}) = v(\overline{\alpha})$. Also,

$$v(\alpha) \leq v_{\overline{y}}(\alpha)$$

for all $\alpha \in A \cap N_{\delta^1}(\overline{\alpha})$ since y fixed restricts the feasible domain.
Consequently, for all $\alpha \in A \cap N_{\delta^1}(\overline{\alpha})$,

$$v(\alpha) < v(\overline{\alpha}) + \varepsilon$$

so that $v(\cdot)$ is u.s.c. relative to A at $\overline{\alpha}$. ||

Condition (a) means that $\overline{y}$ remains feasible at all acceptable
values of α in "close" proximity to $\overline{\alpha}$, that is, at all $\alpha \in A$ or $\hat{A}$ in
some (small) neighborhood of $\overline{\alpha}$. Thus, Theorem 4.1 says that if $\overline{\alpha}$ can
be slightly perturbed in any acceptable fashion without rendering $\overline{y}$
infeasible and if the continuous-variable subproblem $v_{\overline{y}}$ is continuous
at $\overline{\alpha}$, then v is continuous (relative to A) at $\overline{\alpha}$ (since v is l.s.c. by

64

Theorem 3.6).

The interiority condition (a) of Theorem 4.1 is not always necessary. This is shown by the problem

$$\text{minimize } 0$$
$$y \in \{0,1\}$$

$$\text{subject to: } \alpha - 1 \le y \le \alpha \ .$$

The feasible parameter set $A = [0,2]$. Let $\bar{\alpha} = 1$. Both $\bar{y} = 0$ and $\bar{y} = 1$ are optimal. For $\bar{y} = 0$, $A_0 = [0,1]$ so that $pi(A_0|A) = [0,1)$. For $\bar{y} = 1$, $A_1 = [1,2]$ so that $pi(A_1|A) = (1,2]$. Thus, $\bar{\alpha} = 1 \notin pi(A_{\bar{y}}|A)$ for any $\bar{y}$ that is optimal at $\bar{\alpha}$. Yet, v is obviously continuous at $\bar{\alpha} = 1$.

Condition (a) could be replaced by the stronger condition that $\bar{\alpha} \in int(A_{\bar{y}}|A)$. Also, (a) could be replaced by the still stronger condition that $\bar{\alpha} \in int(A_{\bar{y}})$, which is meaningful when $A_{\bar{y}}$ is of full dimension. (In application, it seems that $A_{\bar{y}}$ is more likely to be of full dimension than not.) The conclusion is then also slightly stronger: v is u.s.c. relative to E^r at $\bar{\alpha}$. This follows since $\bar{\alpha} \in int(A)$. We will refer to these alternate conditions as (a)' and (a)":

$$\begin{array}{lll}
\text{(a)'} & & \bar{\alpha} \in int(A_{\bar{y}}|A) \\
\text{(a)''} & & \bar{\alpha} \in int(A_{\bar{y}}) \ . & (4.8)
\end{array}$$

Notice that $int(A_{\bar{y}}) \subseteq int(A_{\bar{y}}|A) \subseteq pi(A_{\bar{y}}|A)$ so that (a)'' $\Rightarrow$ (a)' $\Rightarrow$ (a).

Combining Theorems 3.10 and 4.1 we obtain a significant and general continuity result.

Theorem 4.2: Let f and H be continuous and G l.s.c. on $\bar{\alpha} \times F_{\bar{y}}(\bar{\alpha})$ @ $\bar{y}$, where $\bar{y}$ is optimal at $\bar{\alpha}$. Then (a) and (b) imply that v is continuous relative to A at $\bar{\alpha}$; (a)' and (b) imply that v is continuous at $\bar{\alpha}$.

It may seem at first glance that the interiority conditions are

too nebulous and awkward to be tractable in practical applications. On the contrary, we suggest that these conditions can be assessed usually without excessive difficulty either by direct characterization of the set A_y^- or by an analysis of the problem constraints at an optimal solution, systematized for special problem classes.

In the following example we apply Theorem 4.2 to derive a simple but significant test for continuity in the capacitated facility location problem with nonlinear costs.

Example (The Capacitated Facility Location Problem with Nonlinear Costs):

$$
\begin{array}{ll}
\text{Minimize} & \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij}(x_{ij}) + \sum_{i=1}^{n} d_i y_i \\
x \geq 0 &
\end{array}
$$

$$
\begin{array}{ll}
\text{subject to:} & \sum_{i=1}^{n} x_{ij} = D_j , \quad j = 1,2,\ldots,m \\
\end{array}
$$

(4.9)

$$
\sum_{j=1}^{m} x_{ij} \leq S_i y_i , \quad i = 1,2,\ldots,n
$$

$$
y_i = 0 \text{ or } 1 , \quad i = 1,2,\ldots,n .
$$

The transportation cost from facility i to demand area j, $c_{ij}(\cdot)$, is taken to be a finite and convex function of the quantity x_{ij} on $[0,D_j]$. The vectors D and S denote demand and capacity, respectively, and d the setup costs. Continuity with respect to c is interpreted to mean continuity with respect to any parameters involved in the definitions of the convex functions, c_{ij}. For instance, if $c_{ij} = a_{ij}x_{ij} + b_{ij}x_{ij}^2$ (where $b_{ij} \geq 0$) then v continuous with respect to c means that v is continuous with respect to all parameters a_{ij} and b_{ij}.

We shall address only the parameters D and S since by Theorem 3.6 and Theorem 3.4, v is continuous with respect to all parameters,

66

(c,d,D,S), if v is continuous with respect to D and S. There is a feasible solution for any values of D and S satisfying $D \geq 0$ and $\sum_j D_j \leq \sum_i S_i$. Evidently,

$$A = \{(D,S) \,|\, D \geq 0, \ \sum D_j \leq \sum S_i\} \qquad (4.10)$$

and when y is held fixed at $\overline{y}$, we have

$$A_{\overline{y}} = \{(D,S) \,|\, D \geq 0, \ \sum D_j \leq \sum S_i \overline{y}_i\} . \qquad (4.11)$$

It is immediately apparent that (D,S) is in the set $pi(A_{\overline{y}}|A)$ if total demand is strictly less than total capacity; that is

$$pi(A_{\overline{y}}|A) = \{(D,S) \,|\, D \geq 0, \ \sum D_j < \sum S_i \overline{y}_i\} . \qquad (4.12)$$

(It happens that in this example A and $A_{\overline{y}}$ are of full dimension so that $int(A_{\overline{y}}) = int(A_{\overline{y}}|A) = \{(D,S) \,|\, D > 0, \ \sum D_j < \sum S_i y_i\}$, but this fact is inconsequential to the way in which $pi(A_{\overline{y}}|A)$ is determined; it is only necessary to observe that A and $A_{\overline{y}}$ have the same dimension).

When y is held fixed at $\overline{y}$ the resulting subproblem is a convex program (in the continuous variables x) with linear constraints if each $c_{ij}(\cdot)$ is convex. Also, the set $A_{\overline{y}}$ given by (4.11) and the set $X \equiv \{x \in E^{mn} \,|\, x \geq 0\}$ are then clearly convex. It follows from Theorem 2, Geoffrion [11], that $v_{\overline{y}}$ is convex on $A_{\overline{y}}$. Consequently, $v_{\overline{y}}$ is continuous on $int(A_{\overline{y}})$, i.e., $v_{\overline{y}}$ is continuous at any (D,S) satisfying $D > 0$ and $\sum_j D_j < \sum_i S_i \overline{y}_i$. These conditions, therefore, imply both conditions (a) and (b) of Theorem 4.2.

We may now state the following continuity result from Theorem 4.2 for problem (4.9) with $(c,d,D,S) = (\overline{c},\overline{d},\overline{D},\overline{S})$:

If $\overline{D}$ and $\overline{S}$ satisfy $\overline{D} > 0$ and $\sum_{j=1}^{m} \overline{D}_j < \sum_{i=1}^{n} \overline{S}_i \overline{y}_i$ where $\overline{y}$ is optimal at $(\overline{c},\overline{d},\overline{D},\overline{S})$, then $v(c,d,D,S)$ is continuous at $(\overline{c},\overline{d},\overline{D},\overline{S})$.

67

It is also clear that v is continuous at $(\overline{c},\overline{d},\overline{D},S)$ where S is any vector of capacities satisfying the following:

$$\sum_{j=1}^{m} \overline{x}_{ij} < S_i \leq \overline{S}_i, \quad i \ni \overline{y}_i = 1$$

$$0 \leq S_i \leq \overline{S}_i, \quad i \ni \overline{y}_i = 0$$

where $(\overline{x},\overline{y})$ are optimal at $(\overline{c},\overline{d},\overline{D},\overline{S})$. This follows since such values for S tend to shrink the feasible region, thereby requiring that $v(\alpha) \geq v(\overline{\alpha})$; yet $(\overline{x},\overline{y})$ remain feasible, hence, optimal.

It is tempting to conjecture that v remains continuous for increases in $\overline{D}$ and decreases in $\overline{S}$ so long as total supply at $\overline{y}$ exceeds total demand (i.e., that v is continuous at $(\overline{c},\overline{d},D,S)$ where $D \geq \overline{D}$, $S \leq \overline{S}$ and $\Sigma\, D_j < \Sigma\, S_i \overline{y}_i$). Unfortunately, this conjecture is false; the variables y may not be optimal at such perturbed values for D,S and it is possible that v can be discontinuous at $(\overline{c},\overline{d},D,S)$.

If we allowed the case where $\overline{D} \geq 0$ in the above statement, then we could conclude only that v is continuous relative to A. However, it is obvious that any demand node for which $D_j = 0$ can be eliminated from the network, so we may as well assume that $D_j > 0$ for all j. Therefore, to determine continuity of v, we need only to observe whether at $\overline{y}$ the total capacity exceeds total demand. In the uncapacitated problem continuity obtains provided that $D > 0$. The result extends easily to the problem with both minimum and maximum capacity constraints.

As a numerical illustration, consider the following problem with 3 potential facilities and 5 demand areas:

$$\text{Minimize } \overline{c} \sum_{i=1}^{3} \sum_{j=1}^{5} x_{ij} + \overline{d} \sum_{i=1}^{3} y_i$$

$$\text{subject to: } x_{11} + x_{21} + x_{31} = 4.2$$

$$x_{12} + x_{22} + x_{32} = 4.6$$

$$x_{13} + x_{23} + x_{33} = 3.6$$

$$x_{14} + x_{24} + x_{34} = 3.3 \qquad\qquad (4.13)$$

$$x_{15} + x_{25} + x_{35} = 2.6$$

$$x_{11} + x_{12} + x_{13} + x_{14} + x_{15} \leq 9.5y_1$$

$$x_{21} + x_{22} + x_{23} + x_{24} + x_{25} \leq 9.5y_2$$

$$x_{31} + x_{32} + x_{33} + x_{34} + x_{35} \leq 9.5y_3$$

$$x_{ij} \geq 0, \ y_{ij} = 0 \text{ or } 1; \ i = 1,2,3; \ j = 1,2,3,4,5$$

where $\overline{c}$ and $\overline{d}$ are any positive scalars. By inspection we see that at least two facilities must be opened; since all arc costs are equal it would not be optimal to open all three facilities. Thus, $\overline{y} = (1,1,0)$, $(1,0,1)$, and $(0,1,1)$ are all optimal. Let us look at $\overline{y} = (1,0,1)$. Then, $\overline{x}_{11} = 4.2$, $\overline{x}_{12} = 4.6$, $\overline{x}_{33} = 3.6$, $\overline{x}_{34} = 3.3$, $\overline{x}_{35} = 2.6$, and all other x-variables are zero. Thus,

$$\sum_{j=1}^{5} \overline{x}_{1j} = 4.2 + 4.6 = 8.8 < 9.5\overline{y}_1$$

$$\sum_{j=1}^{5} \overline{x}_{2j} = 0 = 9.5\overline{y}_2$$

$$\sum_{j=1}^{5} \overline{x}_{3j} = 3.6 + 3.3 + 2.6 = 9.5 = 9.5\overline{y}_3$$

so that the third capacity constraint (as well as the demand constraints) is binding. Yet, v is continuous since $\sum D_j = 18.3 < 19 = \sum S_i \overline{y}_i$. Intuitively, v is continuous since any small change in D or S can be absorbed by the nonbinding capacity constraint(s).

69

Specializing a comment made earlier, it is clear that v is also continuous when S_1 is allowed to take any value in the interval $(8.8, 9.5]$ and S_2 is allowed any value in $[0,9.5]$. That is, v is continuous at $(\overline{c},\overline{d},\overline{D},S)$ where $\overline{D} = (4.2,4.6,3.6,3.3,2.6)$, $S_1 \in (8.8,9.5]$, $S_2 \in [0,9.5]$, and $S_3 = \overline{S}_3 = 9.5$.

We learn from this example the significance of Theorem 4.2 when applied to entire classes of problems. Theorem 4.2 is no less effective when applied to a specific problem at hand, but we are less dismayed with a tedious determination of $pi(A_{\overline{y}}|A)$ if the consequent continuity result applies in general to a class of problems, for such a result is then useful when applied to perturbations and/or modifications to the current problem.

Reference to the continuous-variable subproblems of problem (3.9) and continuity of v_y does not rule out the pure integer case. In a pure integer problem, v_y is constant, hence continuous. Of course, if all the data are constrained to be integral as well, then the question of continuity is not pertinent. The class of pure integer problems (with some or all of its data nonintegral) poses a special case of the problem class addressed in the following corollary to Theorem 4.2.

Corollary 4.2.1: Let the objective function in problem (3.9) have the form $f(\cdot,y)$. Suppose f and H are continuous and G l.s.c. on $\overline{\alpha} \times F_{\overline{y}}(\overline{\alpha})$ a y where y is optimal at $\overline{\alpha}$. Then v is continuous relative to A at $\overline{\alpha}$ if $\alpha \in pi(A_{\overline{y}}|A)$; v is continuous at $\overline{\alpha}$ if $\overline{\alpha} \in int(A_{\overline{y}})$.

Proof: When y is fixed at $\overline{y}$, $f(\alpha;\overline{y})$ is a constant, hence $v_{\overline{y}}$ is continuous. ||

A general result that is computationally more practical is obtained

by extending Theorem 4.2 to ϵ-optimal solutions and ϵ-continuous sub-problems.

Theorem 4.3: Let $\bar{y}$ be ϵ_1-optimal at $\bar{\alpha}$ and let f and H be continuous and G l.s.c. on $\bar{\alpha} \times F_y(\bar{\alpha})$ @ $\bar{y}$. Suppose also that $v_{\bar{y}}$ is finite and ϵ_2-continuous relative to $A_{\bar{y}}$ at $\bar{\alpha}$. Then v is $(\epsilon_1+\epsilon_2)$-continuous relative to A at $\bar{\alpha}$ if $\bar{\alpha} \in pi(A_{\bar{y}}|A)$; v is $(\epsilon_1+\epsilon_2)$-continuous at $\bar{\alpha}$ if $\bar{\alpha} \in int(A_{\bar{y}})$.

Proof: By Theorem 3.10, v is l.s.c. at $\bar{\alpha}$ so that for $\gamma > 0$ there exists a $\delta_1 > 0$ such that $\alpha \in A \cap N_{\delta_1}(\bar{\alpha})$ implies

$$v(\bar{\alpha}) - \gamma \le v(\alpha) . \tag{4.14}$$

The interiority condition means that there exists a $\delta_2 > 0$ such that $\alpha \in A \cap N_{\delta_2}(\bar{\alpha})$ implies

$$v(\alpha) \le v_{\bar{y}}(\alpha) < \infty . \tag{4.15}$$

$v_{\bar{y}}$ ϵ_2-continuous at $\bar{\alpha}$ means that for $\gamma > 0$ there exists a $\delta_3 > 0$ such that $\alpha \in A \cap N_{\delta_3}(\bar{\alpha})$ implies

$$v_{\bar{y}}(\alpha) \le v_{\bar{y}}(\bar{\alpha}) + \epsilon_2 + \gamma . \tag{4.16}$$

By ϵ_1-optimality of $\bar{y}$, we have

$$v_{\bar{y}}(\bar{\alpha}) \le v(\bar{\alpha}) + \epsilon_1 . \tag{4.17}$$

Let $\delta = \min(\delta_1, \delta_2, \delta_3)$. By applying (4.14) through (4.17) in succession, we have that for any $\alpha \in A \cap N_{\delta}(\bar{\alpha})$,

$$v(\bar{\alpha}) - \gamma \le v(\bar{\alpha}) + \epsilon_1 + \epsilon_2 + \gamma$$

which means that v is $(\epsilon_1+\epsilon_2)$-continuous relative to A at $\bar{\alpha}$. (Hence, only a Category II discontinuity could occur at $\bar{\alpha}$.) If $\bar{\alpha} \in int(A_{\bar{y}})$, then $A_{\bar{y}}$ and A are of full dimension and it follows that $A \cap N_{\delta}(\bar{\alpha}) = N_{\delta}(\bar{\alpha})$ so that v is continuous relative to E^r at $\bar{\alpha}$. $\quad||$

Theorems 4.1, 4.2, and 4.3 presuppose that an optimal solution vector $\bar{y}$ is available. For some problem classes it is apparent that v_y is continuous for all $y \in Y(\bar{\alpha})$. In this case, an "a priori" conclusion can be made; i.e., we may conclude (subject to interiority) that v is continuous without knowledge of an optimal solution. This observation is stated as follows:

Corollary 4.3.1: Let f and H be continuous and G l.s.c. on $\bar{\alpha} \times X$ @ $Y(\bar{\alpha})$. If v_y is ε-continuous for each $y \in Y(\bar{\alpha})$ and $\bar{\alpha} \in pi(A_y|A)$ for each $y \in Y(\bar{\alpha})$, then v is ε-continuous relative to A at $\bar{\alpha}$.

A similar consequence of Theorem 4.3 addresses continuity on an open subset of A. Define $Y(N) = \underset{\alpha \in N}{\cup} Y(\alpha)$, where N is an open subset of A.

Corollary 4.3.2: Let N be an open subset of A. Let f and H be continuous and G l.s.c. on $N \times X$ @ $Y(N)$. If each y which is feasible for some $\alpha \in N$ is feasible for all $\alpha \in N$ (i.e., if $y \in Y(N) \Rightarrow y \in Y(\alpha)$, $\alpha \in N$) and if v_y is ε-continuous on N, then v is ε-continuous on N.

Proof: y feasible at each $\alpha \in N$ implies that $N \subseteq A_y$. Thus, for any $\alpha \in N$ and $y \in Y(\alpha)$, we have $\alpha \in int(A_y)$ and v_y ε-continuous at α so that v is ε-continuous at α by Corollary 4.3.1. Consequently, v is ε-continuous relative to E^r on N (since N is open).

Example (The Dynamic Lot-Size Model with Nonlinear Costs): Wagner [20], p. 303, discusses the dynamic lot-size model with concave costs. An MIP formulation of this problem where the production and holding costs are permitted to be any continuous functions of the amounts produced is given as follows:

72

$$\text{Minimize} \quad \overset{n}{\underset{i=1}{\Sigma}} \ [c_i(x_i) + d_i y_i]$$
$$x \geq 0$$

$$\text{subject to:} \quad \overset{i}{\underset{j=1}{\Sigma}} x_j \geq \overset{i}{\underset{j=1}{\Sigma}} D_j \ , \quad i = 1, 2, \ldots, n$$

$$x_i \leq M y_i \ , \quad i = 1, 2, \ldots, n \quad (4.18)$$

$$y_i = 0 \text{ or } 1 \ , \quad i = 1, 2, \ldots, n$$

where M is a very large number. It will be shown via Corollary 4.3.2 that v is continuous on A. We may assume that $D_j > 0$ for all j since otherwise the model could be reformulated to eliminate all periods with $D_j = 0$. This model presumes (without loss of generality) that initial inventory is zero. Consequently, $y_1 = 1$ in any feasible solution at $D \ \epsilon \ \overset{n}{\underset{j=1}{\Pi}} \ (0,\infty) \equiv N$. That is, any y that is feasible for some $D \ \epsilon \ N$ is feasible for all $D \ \epsilon \ N$, for one could always take $x_1 = \overset{n}{\underset{j=1}{\Sigma}} D_j$. Secondly, any perturbation in D can be absorbed by x_1: replace x_1 by $x_1 = x_1 + \overset{n}{\underset{j=1}{\Sigma}} (\delta D)_j$ where δD is a perturbation to D. The corresponding change in v is no more than $|c_1(x_1') - c_1(x_1)|$ which is small if δD is small since c_1 is continuous. Clearly, v_y is continuous on N. We conclude via Corollary 4.3.2 that v is continuous with respect to D on N. Finally, letting C denote the set of acceptable values for the parameters defining the $c_i(\cdot)$, and observing that N is also a domain for the setup costs, d, we may state that $v(c,d,D)$ is continuous on $\text{int}(C) \times N \times N$. To allow for zero setup costs, we can take

$$\hat{A} = C \times \overset{n}{\underset{i=1}{\Pi}} \ [0,\infty) \times \overset{n}{\underset{i=1}{\Pi}} \ (0,\infty) \ .$$

Then v for problem (4.18) is continuous relative to $\hat{A}$ at all $(c,d,D) \ \epsilon \ \hat{A}$.

73

In view of Theorem 4.3, we are interested in specific conditions for special problem classes which assure that the optimal value function of a continuous-variable subproblem is continuous. Such conditions coupled with interiority then imply continuity in the MIP problem.

4.3 Strong Feasibility

A concept related to the interiority condition on α is the condition that there be a point in variable space for which all inequality constraints are nonbinding. We will assume in this subsection that H is independent of α and continuous on X @ Y.

Definition: $\overline{y}$ is strongly feasible (s.f.) at $\overline{\alpha}$ if $F_{\overline{y}}(\overline{\alpha})$ has an interior relative to $\{x \in X \mid H(x,\overline{y}) = 0\}$. Equivalently, $\overline{y}$ is s.f. at $\overline{\alpha}$ if $p_{\overline{y}}(\overline{\alpha}) \equiv \{x \in X \mid H(x,\overline{y}) = 0, \ G(\alpha;x,\overline{y}) < 0\} \neq \emptyset$.

Strong feasibility precludes equality constraints that involve α. Equality constraints that do not depend on α could be assumed away as well since by the Implicit Function Theorem the problem may be regarded as one with fewer variables and only inequality constraints (provided that H is continuously differentiable and has a nonsingular Jacobian).

Lemma 4.4: Let G be continuous on $N(\alpha) \times X$ @ y where $N(\alpha)$ is any open set in E^r containing α (H = H(x,y) is continuous on $X(\alpha)$ @ y). Then y s.f. at $\alpha <\ > p_y(\alpha) \neq \emptyset \Rightarrow \alpha \in \text{int}(A_y)$.

Proof: The equivalence is by definition. If y is s.f. at α, then there is an $x \in X$ such that $G(\alpha;x,y) < 0$. Continuity of G with respect to α implies the existence of a neighborhood $N^1(\alpha)$ such that $G(\alpha^1;x,y) \leq 0$ for all $\alpha^1 \in N^1(\alpha)$. Therefore, $N^1(\alpha) \subseteq A_y$ and $\alpha \in \text{int}(A_y)$. $\quad ||$

Strong feasibility is a stronger condition than parameter set

74

interiority. This is illustrated by the facility location problem discussed earlier (example to Theorem 4.2). Any $y_i = 0$ forces $x_{ij} = 0$ so that x cannot be in the interior of $F_y(\cdot)$ for any y with a zero component. The condition $\alpha \in \text{int}(A_y)$, however, holds under mild conditions on the parameters, as discussed in the example.

Although strong feasibility is a more severe condition, in some situations it may be more readily apparent than is parameter set interiority. If strong feasibility holds, then by Lemma 4.4 there is no need to test for interiority in A_y. Suppose that $\overline{y}$ is ε_1-optimal at $\overline{\alpha}$ in problem (3.9). If it is known that $v_{\overline{y}}$ is ε_2-continuous relative to $A_{\overline{y}}$ at $\overline{\alpha}$, then $G(\overline{\alpha};x,\overline{y}) < 0$ for some $x \in F_{\overline{y}}(\overline{\alpha})$ implies that v is $(\varepsilon_1+\varepsilon_2)$-continuous at $\overline{\alpha}$ by Theorem 4.3. (Of course, α-interiority may hold even when equality is forced in some or all of the components of G.)

A special case in which strong feasibility and parameter set interiority are equivalent (with a qualification as in Theorem 4.5 below) is when G has the form $G(\alpha;x,y) = \tilde{G}(x,y) - \alpha$, where G, $\tilde{G}$ and α have the same number of components. But we shall explore a larger class of functions for which these two conditions are equivalent. Let $\alpha = (a,b)$ where b has dimension m and a has dimension r-m. Suppose that the constraints have the form

$$G(a;x,y) \le b . \qquad (4.19)$$

Let A_y^1 denote the projection of A_y onto b-space with $a = \overline{a}$; i.e.,
$$A_y^1 \quad \{b \in E^m | G(\overline{a};x,y) \le b \text{ for some } x \in F_y(\alpha) \} .$$

Theorem 4.5: If G is continuous on $\overline{a} \times F_y(\overline{\alpha})$ @ y and must satisfy (4.19), and if no component of G satisfies $g_i(\overline{a};x,y) = b_i$ for all $x \in X$, then the following conditions are all equivalent:

(a) $\bar{b} \in \text{int}(A_y^1)$

(b) $\bar{\alpha} \in \text{int}(A_y)$

(c) y s.f. at $\bar{\alpha}$.

Proof: Obviously (b) $\Rightarrow$ (a). (c) $\Rightarrow$ (b) by Lemma 4.4. By (a) there

exists a $b < \bar{b}$ such that y is feasible at b; i.e., there exists an

$x \in F_y(\bar{\alpha})$ such that $G(\bar{\alpha};x,y) \leq b < \bar{b}$. Thus, y is s.f. at $\bar{\alpha}$ so that

(a) $\Rightarrow$ (c). $||$

Theorem 4.5 implies that with the given hypotheses, if v is contin-

uous with respect to the right-side data b, then v is continuous with

respect to the other parameters a as well (provided G is continuous as

specified); secondly, for this class of problems, α-interiority and

strong feasibility are equivalent conditions.

Constraints of the form (4.19) are standard in linear MIP problems

and are typical in most other classes of mathematical programming prob-

lems. It was shown in paragraph 3.3 that when the constraints have

this form, directional continuity of v with respect to b implies conti-

nuity with respect to α.

The two types of interiority conditions that we have addressed

both imply that y remains feasible for small perturbations in α. We

are next concerned with finding conditions which imply that y, optimal

at α, remains optimal for certain small perturbations in α, for this

implies continuity of v on a set rather than at a point. Introductory

results on this subject are given in the next two theorems.

Theorem 4.6: Define $S(\bar{\alpha}) = \{\alpha \in A \mid F(\alpha) \subseteq F(\bar{\alpha})\}$. Let G be continuous

on $N(\bar{\alpha}) \times X \otimes \bar{y}$ where $\bar{y}$ is optimal at $\bar{\alpha}$ and $N(\bar{\alpha})$ is any open set in E^r

containing $\bar{\alpha}$. If y is s.f. at $\bar{\alpha}$, then there exists a neighborhood

76

$N^1(\bar{\alpha}) \subseteq N(\bar{\alpha})$ such that $\bar{y}$ is optimal on $N^1(\bar{\alpha}) \cap S(\bar{\alpha})$.

Proof: There exists an $x \in X$ such that $G(\bar{\alpha};x,\bar{y}) < 0$ ($H(x,\bar{y}) = 0$). By continuity of G, there is a neighborhood $N^1(\bar{\alpha}) \subseteq N(\bar{\alpha})$ such that $\alpha \in N^1(\bar{\alpha}) \rightarrow G(\alpha;x,\bar{y}) < 0$. Hence, $(x,\bar{y})$ remains feasible on $N^1(\bar{\alpha})$ so that f independent of α implies $v(\alpha) \leq v(\bar{\alpha})$ for all $\alpha \in N^1(\bar{\alpha})$. Consequently, $v(\alpha) = v(\bar{\alpha})$ on $N^1(\bar{\alpha}) \cap S(\bar{\alpha})$ so that $\bar{y}$ must be optimal on this set. $||$

Theorem 4.7: Let f be continuous on $\bar{\alpha} \times X(\bar{\alpha})$ @ $\bar{y}$ where $\bar{y}$ is optimal at $\bar{\alpha}$, and let G be continuous on $N(\bar{\alpha}) \times X$ @ $\bar{y}$. If $\bar{y}$ is s.f. at $\bar{\alpha}$, then there exists a neighborhood $N_\varepsilon(\bar{\alpha})$ on which $\bar{y}$ is ε-optimal.

Proof: v is l.s.c. at $\bar{\alpha}$ by Theorem 3.10; hence, there is a neighborhood $N_\varepsilon^1(\bar{\alpha})$ on which $v(\bar{\alpha}) < v(\alpha) + \varepsilon$. By Theorem 4.6 there is a $N^1(\bar{\alpha})$ on which $v(\alpha) \leq v(\bar{\alpha})$ (see proof). Let $N_\varepsilon(\bar{\alpha}) = N_\varepsilon^1(\bar{\alpha}) \cap N^1(\bar{\alpha})$. Then for all $\alpha \in N_\varepsilon(\bar{\alpha})$, $\bar{y}$ is feasible and

$$v(\alpha) \leq v(\bar{\alpha}) < v(\alpha) + \varepsilon .$$

It follows that $\bar{y}$ is ε-optimal on $N_\varepsilon(\bar{\alpha})$. $||$

CHAPTER 5

CONTINUITY PROPERTIES OF CONVEX MIP PROBLEMS

An MIP problem where the objective function cr one or more of the
constraint functions satisfies some type of convexity (or concavity) is
loosely referred to here as a convex problem. The optimal value func-
tion v generally will not be convex or concave. In this chapter we con-
centrate on the particular convexity assumptions that assure continuity
of $v(\cdot)$.

5.1 Convexity Assumptions

The set X is assumed to be convex as well as bounded. Let the data
be represented as $a = (c,a,b,d)$ and rewrite problem (3.9) in the form

$$\begin{cases} \text{Minimize } f(c;x,y) \\ (x,y)\epsilon U \\ \text{subject to:} \quad G(a;x,y) \leq b \\ \qquad\qquad\qquad H(a;x,y) = d \end{cases} \tag{5.1}$$

where $U \supseteq X \times Y$. The set U is employed to accommodate constraints in
both x and y that do not involve data "of interest."

Problem (5.1) is actually no less general than problem (3.9): it
is not required that c, a, b, and d be independent or that b and d be
something other than zero-vectors (that is, problem (3.9) results by
setting b and d to zero and replacing c and a by a). However, it will
be instructive to view b and d as scarce resources.

As before, the following notation will be used:

$$\bar{v}(c) \quad v(c,\bar{a},\bar{b},\bar{d}) \text{ where } \bar{a},\bar{b},\bar{d} \text{ are fixed}$$

$$\bar{v}(b,d) \quad v(c,a,b,d) \text{ where } \bar{c},\bar{a} \text{ are fixed }.$$

78

Similar definitions hold for $v_y(\cdot)$. Define also

$$\overline{Z}_y \equiv \{(b,d) \,|\, F_y(\overline{c},\overline{a},b,d) \neq \emptyset\} \tag{5.2}$$

where $F_y(\overline{c},\overline{a},b,d)$ denotes the feasible region in (5.1) when y, $\overline{c}$, and $\overline{a}$ are held fixed.

Theorem 5.1: Given that f and G are convex and H linear in x for y and α fixed, and that f is continuous on $\overline{c} \times X @ Y$ and G and H are continuous on $\overline{a} \times X @ Y$:

(a) $\overline{Z}_y$ is a convex set, $y \in Y$

(b) $\overline{v}_y(b,d)$ is continuous and convex on $\overline{Z}_y$, $y \in Y$

(c) v is ε-continuous at $\overline{\alpha}$ if $(\overline{b},\overline{d}) \in \text{int}(\overline{Z}_y)$ for some y that is ε-optimal at $\overline{\alpha}$.[*]

Proof: (a) Let $(b_1,d_1),(b_2,d_2) \in \overline{Z}_y$, $y \in Y$. By definition, there exist $x_1, x_2 \in X$ such that

$$G(a;x_1,y) \leq b_1 , \qquad G(a;x_2,y) \leq b_2$$
$$H(a;x_1,y) = d_1 , \qquad H(a;x_2,y) = d_2 .$$

We see that $(b,d) \equiv \lambda(b_1,d_1) + (1-\lambda)(b_2,d_2) \in \overline{Z}_y$, where $\lambda \in [0,1]$; this follows since $x \equiv \lambda x_1 + (1-\lambda)x_2$ satisfies:

$$G(a;x,y) \leq \lambda G(a;x_1,y) + (1-\lambda)G(a;x_2,y)$$
$$\leq \lambda b_1 + (1-\lambda)b_2$$
$$\qquad b$$

$$H(a;x,y) = \lambda H(a;x_1,y) + (1-\lambda)H(a;x_2,y)$$
$$= \lambda d_1 + (1-\lambda)d_2$$
$$\qquad d .$$

---

[*] Result (c) holds without the convexity assumptions on f and G.

79

(b) $\bar{v}_y(b,d)$ is l.s.c. on $\bar{Z}_y$ by Theorem 3.10. It suffices to show that $\bar{v}_y(b,d)$ is convex on $\bar{Z}_y$ (this can be deduced from Geoffrion [11], Theorem 2, but we present the proof in the current context). Let (b,d) be defined as in part (a).

$$\bar{v}_y(b,d) = \min_{x_1,x_2 \in X} f(c;\lambda x_1 + (1-\lambda)x_2,y) \left|
\begin{array}{l}
G(a;\lambda x_1 + (1-\lambda)x_2,y) \le b \\[2mm]
H(a;\lambda x_1 + (1-\lambda)x_2,y) = d
\end{array}
\right.$$

$$\le \min_{x_1,x_2 \in X} f(c;\lambda x_1 + (1-\lambda)x_2,y) \left|
\begin{array}{l}
G(a;x_1,y) \le b_1,\ G(a;x_2,y) \le b_2 \\[2mm]
H(a;x_1,y) = d_1,\ H(a;x_2,y) = d_2
\end{array}
\right.$$

$$\le \min_{x_1,x_2 \in X} \lambda f(c;x_1,y) + (1-\lambda)f(c;x_2,y) \left|
\begin{array}{l}
G(a;x_1,y) \le b_1,\ G(a;x_2,y) \le b_2 \\[2mm]
H(a;x_1,y) = d_1,\ H(a;x_2,y) = d_2
\end{array}
\right.$$

$$= \lambda \bar{v}_y(b_1,d_1) + (1-\lambda)\bar{v}_y(b_2,d_2)\ .$$

(c) Suppose that $\bar{y}$ is ε-optimal at $\bar{\alpha} \equiv (\bar{c},\bar{a},\bar{b},\bar{d})$ and that $(\bar{b},\bar{d}) \in \mathrm{int}(\bar{Z}_y)$. If $\bar{\alpha} \in \mathrm{int}(A_y^{\to})$, then the desired conclusion is immediate from Theorem 4.3 since $\bar{v}_y$ is continuous (convex, in fact). By hypothesis, there exists an open ball $N_\delta(\bar{b},\bar{d}) \subseteq \bar{Z}_y$. That is, for any (b,d) such that $\| (b,d)-(\bar{b},\bar{d}) \| < \delta$ there exists an $x \in X$ such that $G(\bar{a};x,\bar{y}) \le b$ and $H(\bar{a};x,\bar{y}) = d$. The hypothesized continuity of G and H imply the existence of a neighborhood $N(\bar{a})$ about $\bar{a}$ such that

$$a \in N(\bar{a}) \implies \left\| \begin{array}{l} G(a;x,\bar{y}) - G(\bar{a};x,\bar{y}) \\[2mm] H(a;x,\bar{y}) - H(\bar{a};x,\bar{y}) \end{array} \right\| < \delta/2 \tag{5.3}$$

where δ is independent of x and $\bar{y}$ since $X \times Y$ is bounded. Let $(a,b,d) \in N(\bar{a}) \times N_{\delta/2}(\bar{b},\bar{d})$ and write the constraints

$$G(a;x,\bar{y}) \le b, \qquad H(a;x,\bar{y}) = d \tag{5.4}$$

in the form

$$G(\overline{a};x,\overline{y}) \leq \check{b} \equiv b + G(\overline{a};x,\overline{y}) - G(a;x,\overline{y})$$

$$H(\overline{z};x,\overline{y}) = \check{d} \equiv d + H(\overline{a};x,\overline{y}) - H(a;x,\overline{y}) \ . \tag{5.5}$$

Now $(\check{b},\check{d}) \ \varepsilon \ N_{\delta}(\overline{b},\overline{d})$ since, using (5.3) and the fact that $(b,d) \ \varepsilon$
$N_{\delta/2}(\overline{b},\overline{d})$,

$$\left\| \begin{pmatrix} \check{b} \\ \check{d} \end{pmatrix} \right\| = \left\| \begin{pmatrix} b-\overline{b} \\ d-\overline{d} \end{pmatrix} + \begin{bmatrix} G(\overline{a};x,\overline{y}) - G(a;x,\overline{y}) \\ H(\overline{a};x,\overline{y}) - H(a;x,\overline{y}) \end{bmatrix} \right\|$$

$$\leq \left\| \begin{pmatrix} b-\overline{b} \\ d-\overline{d} \end{pmatrix} \right\| + \left\| \begin{matrix} G(\overline{a};x,\overline{y}) - G(a;x,\overline{y}) \\ H(\overline{a};x,\overline{y}) - H(a;x,\overline{y}) \end{matrix} \right\|$$

$$< \delta/2 + \delta/2 = \delta \ .$$

This means there exists an $x' \ \varepsilon \ X$ satisfying (5.5), hence (5.4); there-
fore $N(\overline{a}) \times N_{\delta/2}(\overline{b},\overline{d})$ is contained in the set of feasible values for
(a,b,d) and consequently v is continuous with respect to (a,b,d) by
Theorem 4.3. It follows from Theorem 3.6 that v is continuous with
respect to $\alpha \equiv (c,a,b,d)$. (Notice that result (c) holds without the
convexity assumptions on f and G -- only continuity is needed.) $||$

This result states that (under the hypotheses) if small perturba-
tions in $(\overline{b},\overline{d})$ do not destroy feasibility when $\overline{y}$ (optimal at $\overline{\alpha}$) is fixed,
then v is continuous at $\overline{\alpha}$. The computational leverage that this result
provides in a continuity analysis is that one needs only to address the
right-side parameters (b,d). Notice that conclusion (c) implies that
$v(\cdot)$ is discontinuous at $\overline{\alpha}$ if and only if $\overline{v}(b,d)$ is discontinuous at
$(\overline{b},\overline{d})$.

With regard to problem (3.9) wherein the right sides of the con-
straints are zero, the parameters (b,d) may be viewed as perturbations.
Application of Theorem 5.1 then states that, under the hypotheses, v is
continuous not only with respect to α but also with respect to

81

perturbations (b,d) about zero to the right side.

The interiority condition of Theorem 5.1(c) is equivalent to strong feasibility if (b,d) is regarded as part of the data α (Theorem 4.5).[*] The next result applies to problem (3.9) but the strong feasibility hypothesis can be replaced by the interiority condition of Theorem 5.1(c) in the context of problem (5.1).

Theorem 5.2: Let f and G be continuous on $\bar{\alpha} \times X \oplus Y$. Suppose that G is convex in x for fixed α and y and that H is not dependent on α. Then the existence of a $\bar{y} \in Y$ that is optimal and strongly feasible at $\bar{\alpha}$ implies that $M_\epsilon(\cdot)$ is closed at $\bar{\alpha}$.

Proof: $F_{\bar{y}}(\cdot)$ is open at $\bar{\alpha}$ by Theorem 12, Hogan [16]; hence, $F(\cdot)$ is open at $\bar{\alpha}$ (since $F(\cdot) = \bigcup_{y \in Y} F_y(\cdot)$ and an arbitrary union of open maps is an open map -- see Berge [4]). Also, $F(\cdot)$ is closed by Theorem 3.9; so $F(\cdot)$ is continuous at $\bar{\alpha}$ and the result follows from Theorem 3.5. $||$

The optimal value function v is the pointwise minimum of the finitely many functions v_y: $v(\alpha) = \min_{y \in Y} \{v_y(\alpha)\}$. Under the conditions of Theorem 5.1, each function $\bar{v}_y(b,d)$ is convex so that $\bar{v}(b,d)$ is piecewise convex. Now, in order for $v(\cdot)$ to be discontinuous at a point $\bar{\alpha}$ it must be that $(\bar{b},\bar{d})$ is such that each $\bar{y} \in M_0(\bar{\alpha})$ becomes infeasible for some arbitrarily small perturbation in $(\bar{b},\bar{d})$. That is, for any $\bar{y} \in M_0(\bar{\alpha})$, $(\bar{b},\bar{d}) \in \mathrm{bd}(\bar{Z}_{\bar{y}})$. The next result states that the set of such points is relatively small. Define:

---

[*] If (b,d) is not part of the data, i.e., $\alpha \equiv (c,a)$, then one could have $\bar{\alpha} \in \mathrm{int}(A_y)$ where y is not strongly feasible at $\bar{\alpha}$.

$$\overline{W}_y \equiv \{(b,d) \mid y \text{ is optimal}\} \subseteq \overline{Z}_y$$

$$D \equiv \{(b,d) \mid \overline{v}(b,d) \text{ is discontinuous at } (b,d)\}$$

$\mu(D) \equiv$ (Lebesque) measure of D relative to the smallest subspace containing D.

__Theorem 5.3:__ $\mu(D) = 0$. (The set of points α at which $v(\cdot)$ is discontinuous has relative measure zero).

__Proof:__ By Theorem 5.1(c), $v(\cdot)$ is discontinuous at $\overline{\alpha}$ if and only if $\overline{v}(b,d)$ is discontinuous at $(\overline{b},\overline{d})$. Let t be the smallest dimension such that $D \subseteq E^t$ and let T be an arbitrary subset of E^t with finite measure. Per the above discussion, $D \subseteq \bigcup_{y \in Y} bd(\overline{W}_y)$. Since $\mu[cl(\overline{W}_y)] = \mu[int(\overline{W}_y)]$ where the interior is taken relative to E^t, it follows that

$$\mu(T \cap D) \leq \sum_{y \in Y} \mu[T \cap bd(\overline{W}_y)] = \sum \mu(T \cap [cl(\overline{W}_y) \sim int(\overline{W}_y)])$$

$$= \sum \mu[T \cap cl(\overline{W}_y)] - \sum \mu[T \cap int(\overline{W}_y)]$$

$$= 0 \quad \text{(since each term is finite)}.$$

The result now follows because T is arbitrary. $\|$

__Corollary 5.3.1:__ Given a data point α selected at random from A, the probability that v is discontinuous at α is zero.

5.2 Concavity of v with Respect to Cost Data

Concerned only with objective function data, the problem (i.e., (5.1) or (3.9)) can be expressed as

$$\begin{cases} \text{minimize } f(c;x,y) \\ \text{subject to:} \quad (x,y) \in F(\overline{a}) \ . \end{cases} \tag{5.6}$$

The optimal value function $\overline{v}(c) \equiv v(c,\overline{a})$, $\overline{a}$ fixed, is continuous on E^n where n is the dimension of c, provided that f is continuous on $E^n \times F(\overline{a})$.

It also holds that $\overline{v}(c)$ is concave on E^n if f is concave in c. The function f is concave in c if for any $(x,y) \in F(\overline{a})$ and $\lambda \in [0,1]$, $f(\lambda c_1 + (1-\lambda)c_2; x,y) \geq \lambda f(c_1; x,y) + (1-\lambda)f(c_2; x,y)$. This is a very mild restriction; in particular, f linear in c permits f to represent any polynomial on $X \times Y$.

Theorem 5.4: If f is concave in c, then $\overline{v}(c)$ is concave on E^n.

Proof: Let $c \equiv \lambda c^1 + (1-\lambda)c^2$ where $\lambda \in [0,1]$.

$$\overline{v}(c) = \underset{x \in X, y \in Y}{\text{minimum}} f(c; x,y) \mid (x,y) \in F(a)$$

$$\geq \underset{x \in X, y \in Y}{\text{minimum}} \lambda f(c^1; x,y) + (1-\lambda)f(c^2; x,y) \mid (x,y) \in F(a)$$

$$\geq \underset{x \in X, y \in Y}{\text{minimum}} f(c^1; x,y) \mid (x,y) \in F(a) + \underset{x \in X, y \in Y}{\text{minimum}} (1-\lambda)f(c^2; x,y) \mid (x,y) \in F(a)$$

$$= \lambda \overline{v}(c^1) + (1-\lambda)\overline{v}(c^2) \ . \qquad \qquad ||$$

5.3 Quasiconvexity of Constraint Functions

A function $g(\alpha; x)$ is _quasiconvex_ on $A \times X$ if for $(\alpha_1, x_1) \in A \times X$, $(\alpha_2, x_2) \in A \times X$, and $\lambda \in [0,1]$,

$$g(\lambda \alpha_1 + (1-\lambda)\alpha_2; \lambda x_1 + (1-\lambda)x_2) \leq \max[g(\alpha_1, x_1), g(\alpha_2, x_2)] \ . \qquad (5.7)$$

This definition is extended in the obvious way to a function $g(\alpha; x,y)$ quasiconvex on $A \times X @ y$. The vector function G is quasiconvex if each component function g is quasiconvex. A function g is quasiconcave if $-g$ is quasiconvex.

Lemma 5.5: If N is an open subset of A and $F(\cdot)$ is closed on $cl(N)$, then $cl(N) \subseteq A$ (i.e., $F(\alpha) \neq \emptyset$ for $\alpha \in cl(N)$).

Proof: It follows from Berge [4], Theorem 2 (Chapter 6), that $F(\hat{\alpha}) = \emptyset$ implies the existence of a neighborhood about $\hat{\alpha}$ on which $F(\cdot) = \emptyset$. If

84

$\hat{\alpha} \in S \equiv cl(N)-N$ there can be no such neighborhood since $N \subseteq A$. Hence, $F'(\hat{\alpha}) \neq \emptyset$ and $S \subseteq A$.

Lemma 5.6: Let $N_\delta(\overline{\alpha})$ be a δ-neighborhood of $\overline{\alpha}$ and define $S \equiv cl[N_\delta(\overline{\alpha})] - N_\delta(\overline{\alpha})$. Then for any $\alpha \in N_\delta(\overline{\alpha})$ there is an $\hat{\alpha} \in S$ such that $\alpha = \lambda\hat{\alpha} + (1-\lambda)\overline{\alpha}$ where $\lambda \in [0,1]$.

Proof: For any $\hat{\alpha} \in S$, $\|\overline{\alpha}-\hat{\alpha}\| = \delta$. Let $\hat{\alpha} = \overline{\alpha} + \delta \frac{\alpha-\overline{\alpha}}{\|\alpha-\overline{\alpha}\|}$. Then $\alpha = \lambda\hat{\alpha} + (1-\lambda)\overline{\alpha}$ where $0 < \lambda \equiv \frac{1}{\delta}\|\alpha-\overline{\alpha}\| \leq 1$. ($\lambda = 0$ if and only if $\alpha = \overline{\alpha}$.) $||$

Theorem 5.7: Let N be an arbitrary open subset of $A_{\overline{y}}$. Suppose that:

(a) X is convex (as well as bounded),

(b) G is quasiconvex and l.s.c. on $cl(N) \times X$ @ $\overline{y}$,

(c) each component of H is either independent of α or linear on $cl(N) \times X$ @ $\overline{y}$, and

(d) f is continuous on $cl(N) \times X$ @ $\overline{y}$.

Then $v_{\overline{y}}$ is continuous (and finite) on N and v is continuous (and finite) at all points in N at which $\overline{y}$ is optimal.

Proof: $F_{\overline{y}}(\cdot)$ is closed on $cl(N)$ by Theorem 3.9. Hence, $cl(N) \subseteq A_{\overline{y}}$ by Lemma 5.5. To see that $F_{\overline{y}}(\cdot)$ is open on N, let $\overline{x} \in F_{\overline{y}}(\overline{\alpha})$ where $\overline{\alpha} \in N$ and let $\alpha^k \to \overline{\alpha}$. Since N is open, $\alpha^k \in N$ for k large. By Lemma 5.6 there exist $\hat{\alpha}^k \in S \equiv cl(N)-N$ such that $\alpha^k = \lambda_k\hat{\alpha}^k + (1-\lambda_k)\overline{\alpha}$ where necessarily $\lambda_k \to 0$. Now $S \subseteq A_{\overline{y}}$ so there exists $\hat{x}^k$ such that $G(\hat{\alpha}^k;\hat{x}^k,\overline{y}) \leq 0$ and $H(\hat{\alpha}^k;\hat{x}^k,\overline{y}) = 0$. Let $x^k = \lambda_k\hat{x}^k + (1-\lambda_k)\overline{x} \to \overline{x}$ (since each $\hat{x}^k$ belongs to the bounded set X). By quasiconvexity of G,

$$G(\alpha^k;x^k,y) \leq \max \{G(\hat{\alpha}^k;\hat{x}^k,\overline{y}),G(\overline{\alpha};\overline{x},\overline{y})\} \leq 0 . \qquad (5.8)$$

The component functions, h_i, of H that do not depend on α have no influence on the continuity of $F_{\overline{y}}$, so we may ignore them. The other

85

components satisfy

$$h(\alpha^k;x^k,\overline{y}) = \lambda_k h(\hat{\alpha}^k;\hat{x}^k,\overline{y}) + (1-\lambda_k)h(\overline{\alpha};\overline{x},\overline{y}) = 0 . \qquad (5.9)$$

By (5.8) and (5.9), $x^k \in F_{\overline{y}}(\alpha^k)$ so that $F_{\overline{y}}(\cdot)$ is open at $\overline{\alpha}$. Since $\overline{\alpha}$ is an arbitrary element of the open set N, it follows that $F_{\overline{y}}(\cdot)$ is open on N. Consequently, $F_{\overline{y}}(\cdot)$ is continuous on N and by Theorem 3.4 (with $\overline{y}$ fixed) we conclude that $v_{\overline{y}}$ is continuous on N; $v_{\overline{y}}$ is finite on N since $N \subseteq A_{\overline{y}}$ and f is bounded (continuous on the compact set $cl(N) \times X$). Finally, if $\overline{y}$ is optimal at $\alpha \in N$, then $\alpha \in int(A_{\overline{y}})$ and $v_{\overline{y}}$ is continuous at α so that v is continuous at α by Theorem 4.3. $\qquad ||$

Corollary 5.7.1: Suppose $\overline{y}$ is ε-optimal at $\overline{\alpha} \in pi(\hat{A}_{\overline{y}}|\hat{A})$. Given conditions (a) through (d), v is ε-continuous relative to $\hat{A}$ at $\overline{\alpha}$.

Proof: (Immediate from Theorems 4.3 and 5.7).

The condition that a component h of H be linear on a product set N × X permits the form

$$h(\alpha;x,\overline{y}) = \ell(x,\overline{y}) - \alpha\overline{y} - \alpha$$

where $\ell(x,\overline{y})$ is linear in x for $\overline{y}$ fixed. The form αx is not acceptable in equality constraints since it is not linear on N × X. Quasiconvexity of the inequality functions is also a strong condition; it too precludes the form αx for $\alpha \geq 0$ and $x \geq 0$. In fact, αx is quasiconcave for non-negative α and x.

Remark: αx is quasiconvex for $\alpha \leq 0$ and $x \geq 0$.

Proof: αx is quasiconvex for $\alpha \leq 0$ and $x \geq 0$ if and only if the sets $\{(\alpha,x) | \alpha \geq 0, x \geq 0, \alpha x \geq \gamma\}$ are convex for any $\gamma \geq 0$ (Theorem 2.10, Zangwill [24]). It is sufficient to consider α and x as scalars since

if $C_1, \ldots, C_p$ are convex sets in E^1, then $C_1 \times \ldots \times C_p$ is convex in E^p (Theorem 3.5, Rockafellar [19]). Given $\alpha_1 x_1 \geq \gamma$ and $\alpha_2 x_2 \geq \gamma$ we need to establish that

$$[\lambda\alpha_1 + (1-\lambda)\alpha_2] \cdot [\lambda x_1 + (1-\lambda)x_2] \geq \gamma . \qquad (5.10)$$

If $\alpha_1 = 0$ or $\alpha_2 = 0$, then $\gamma = 0$ and the result is immediate since l.h.s.(5.10) ≥ 0. We will assume that $\alpha_1 > 0$ and $\alpha_2 > 0$. Now,

$$\text{l.h.s.}(5.10) = \lambda^2 \alpha_1 x_1 + (1-\lambda)^2 \alpha_2 x_2 + \lambda(1-\lambda)(\alpha_2 x_1 + \alpha_1 x_2)$$

$$\geq [\lambda^2 + (1-\lambda)^2]\gamma + (\lambda-\lambda^2)(\frac{\alpha_2}{\alpha_1} + \frac{\alpha_1}{\alpha_2})\gamma \geq \gamma$$

since $\dfrac{\alpha_2}{\alpha_1} + \dfrac{\alpha_1}{\alpha_2} \geq 2$. $\qquad\qquad || $

An example of a more general quasiconvex form is

$$q_1(\alpha;\overline{y}) + q_2(x,\overline{y}) + \alpha^{(1)}x + \alpha^{(2)}\overline{y} + \alpha_0$$

where $\overline{y}$ is fixed, $x \geq 0$, $\alpha^{(1)} \leq 0$, q_1 if quasiconvex in α, and q_2 is quasiconvex in x.

5.4 Summary and Conclusions

Two interesting special applications of Theorems 5.1, 5.2, and 5.3 are to LP problems and linear MIP problems. The conclusions in these theorems hold for all bounded linear MIP problems. Applied to LP problems, Theorems 5.1 and 5.3 can be summarized as follows: in any (bounded) LP problem the set $\overline{Z} \equiv \{(b,d) \,|\, -\infty < v(\overline{\alpha}) < \infty\}$ is convex and the optimal value v is (a) a concave function of the objective function data, (b) a convex function of the right-side data, and (c) is continuous at a point $\overline{\alpha}$ if $(\overline{b},\overline{d}) \in \text{int}(\overline{Z})$.

It may appear at first that the convexity assumption on f in

87

Theorem 5.1 and the concavity assumption on f in Theorem 5.3 preclude simultaneous application of these theorems in all but the linear case. Indeed, f can be both concave in c and convex in x. The simple quadratic: $c_1 x_1^2 + c_2 x_2^2$, for example, is linear in c and convex in x for $c_1, c_2 \geq 0$.

In the case of MIP problems with convex objective and constraint functions (linear equality constraint functions), the continuity of v at a point is assured by a single interiority condition on only the right-side data. Viewed as a function only of the right-side data, the optimal value is piecewise convex. This follows since each subproblem (y fixed) has optimal value convex in the right-side data and v is the pointwise minimum of these functions: $v(\alpha) = \min_{y \in Y} \{ v_y(\alpha) \}$. As shown (one-dimensionally) in Figure 5.1, v can be discontinuous at (and only at) right-side points where $y \in Y$ become infeasible. The optimal value function is shown in bold face.
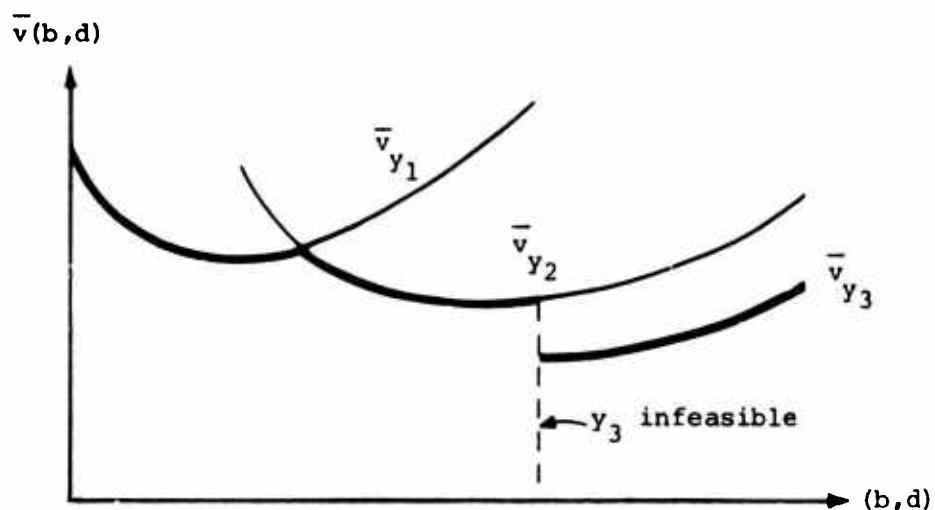
Figure 5.1.   The Optimal Value is Piecewise Convex with Respect to the Right-Side Data.

In the case where the (inequality) constraint functions are quasi-convex in a product space, continuity of v follows without an explicit interiority requirement. However, it was pointed out that quasiconvexity jointly in α and x is a quite severe assumption; the severity is mitigated somewhat in Theorem 5.7 by requiring quasiconvexity only on a neighborhood N of $\bar{\alpha}$. In one respect, the conditions of Theorem 5.7 permit a more general application than do those of Theorem 5.1 in that discontinuous constraint functions are allowed -- a function can be both quasiconvex and lower semicontinuous and yet be discontinuous.

CHAPTER 6

CONTINUITY PROPERTIES OF LINEAR MIP PROBLEMS

The general linear mixed integer programming problem is expressed
as

$$
\begin{cases}
\text{minimize } c'x + c''y \\
(x,y) \epsilon U \\
\text{subject to: } G_1 x + G_2 y \leq \vdash \\
\qquad\qquad\quad H_1 x + H_2 y = d
\end{cases}
\tag{6.1}
$$

where $U \subseteq X \times Y$, X is a bounded polytope in E^p, Y is a finite set in
E^q, $c' \epsilon E^p$, $c'' \epsilon E^q$, b has m_1 components and d has m_2 components; G_1
is an $m_1 \times p$ matrix, G_2 is $m_1 \times q$, H_1 is $m_2 \times p$ and H_2 is $m_2 \times q$. Let
$n = p+q$ and $m = m_1+m_2$. Let $c \epsilon E^n$ denote (c',c'') and define $G \equiv [G_1 : G_2]$,
$H \equiv [H_1 : H_2]$.

The vector α will denote all elements of c, G, H, b, and d. We
will assume that these parameters are constants; however, the results
in this section do not rule out the case where α, i.e., (C,G,H,b,d)
varies continuously with respect to auxiliary parameters β; v is conti-
nuous with respect to β if v is continuous with respect to α. Problem
(6.1) being a special case of a convex problem, it follows from results
of the preceding chapter that v is continuous with respect to α if v is
continuous with respect to b and d (relative to E^m).

6.1 Basic Properties of the Optimal Value Function

Since X and Y are bounded sets, the objective function is finite
for all choices of c' and c''. Consequently, $v(\alpha)$ is finite for all
$\alpha \epsilon A$, where A again denotes the feasible parameter set: $\{\alpha | F(\alpha) \neq \emptyset\}$.
We define $v(\alpha) = \infty$ for $\alpha \notin A$.

One of the most significant continuity properties of linear MIP problems follows immediately from Theorem 3.10:

<u>Theorem 6.1:</u> $v(\cdot)$ is l.s.c. on A.

Thus, a "small" perturbation in α cannot induce a discrete reduction in the optimal value. More precisely, at any point $\alpha \in A$ and for any $\varepsilon > 0$, there exists a neighborhood about α such that the infimal value v at any point in this neighborhood is not less than $v(\alpha)-\varepsilon$.

A property that can be viewed as a companion to Theorem 6.1 is that each set A_y is closed. This is equivalent to the following:

<u>Theorem 6.2:</u> There exists a neighborhood $N(\overline{\alpha})$ about each point $\overline{\alpha}$ such that $Y(\alpha) \subseteq Y(\overline{\alpha})$ for all $\alpha \in N(\overline{\alpha})$.

<u>Proof:</u> If $\overline{\alpha} \notin A$ there exists a $N(\overline{\alpha})$ not in A so that $Y(\alpha) = \emptyset$ for all $\alpha \in N(\overline{\alpha})$. For $\overline{\alpha} \in A$, $F(\cdot)$ is closed at $\overline{\alpha}$ by Theorem 3.9 and the result obtains from Theorem 3.11. $||$

This result states that any value of y that is infeasible at $\overline{\alpha}$ cannot be feasible at any point arbitrarily close to $\overline{\alpha}$. That is, each set A_y is closed. The consequence is that small perturbations in $\overline{\alpha}$ cannot admit "new" y vectors (i.e., y vectors that are not feasible at $\overline{\alpha}$) so that any increment to the feasible domain of problem (6.1) is due to admission of additional values for the continuous variables x. This observation will lend insight into some of our later results.

Several elementary properties of v in problem (6.1) are easily observed. These regard the behavior of v with respect to certain of the parameters when the other parameters are held fixed. To avoid confusion we define precisely what we mean by "with respect to certain parameters."

Letting $\bar{\alpha}$ denote a fixed point, we define the projected functions:

$$\bar{v}(c) \equiv v(c,\bar{G},\bar{H},\bar{b},\bar{d})$$

$$\bar{v}(b,d) \equiv v(\bar{c},\bar{G},\bar{H},b,d)$$

$$\bar{v}(b) \equiv v(\bar{c},\bar{G},\bar{H},b,\bar{d}) \tag{6.2}$$

$$\bar{v}(G,H) \equiv v(\bar{c},G,H,\bar{b},\bar{d}) \ .$$

When y is fixed (and $c"$, G_2, and H_2 are fixed), the subfunctions v_y depend only on c', G_1, H_1, b, and d. In like manner, define

$$\bar{v}_y(c') \equiv v_y(c',\bar{c}",\bar{G},\bar{H},\bar{b},\bar{d})$$

$$\bar{v}_y(b,d) \equiv v_y(\bar{c},\bar{G},\bar{H},b,d)$$

$$\bar{v}_y(b) \equiv v_y(\bar{c},\bar{G},\bar{H},b,\bar{d}) \tag{6.3}$$

$$\bar{v}_y(G_1,H_1) \equiv v_y(\bar{c},G_1,\bar{G}_2,H_1,\bar{H}_2,\bar{b},\bar{d}) \ .$$

Theorem 6.3:

(a) $\bar{v}(c)$ is piecewise linear and concave (hence, continuous) on E^n.

(b) $\bar{v}(c,G,b)$ is continuous at $(\bar{c},\bar{G},\bar{b})$ if $\bar{v}(b)$ is continuous at $\bar{b}$.

(c) $\bar{v}(b)$ is monotone nonincreasing with respect to b.

(d) If $x \geq 0$ and $y \geq 0$ is required, then $\bar{v}(G,H)$ is monotone nondecreasing with respect to G and H, and $\bar{v}(c)$ is monotone nondecreasing.

Proof:

(a) The convex hull of the bounded domain $F(\alpha)$ in problem (6.1) can be characterized by a finite collection S of linear constraints, the extreme points of which satisfy the integer requirements. Problem (6.1) at α and the LP problem: $\min c'x + c"y \,|\, S$ are equivalent in the sense that they possess the same optimal solutions and solution values. The result follows from LP theory since S is invariant with respect to c (see Theorem 6.4(b)).

(b) $\overline{v}(b)$ continuous at $\overline{b}$ implies that $\hat{v}(\theta)$ (see paragraph 3.3) is continuous at 0. The result is immediate from Theorem 3.12.

(c) An increment in b increases the feasible domain; $\overline{v}(b)$ may therefore only decrease in value or remain the same.

(d) An increment in G reduces the size of the feasible domain so $\overline{v}(G)$ may only increase in value or remain the same. ||

6.2 Continuity Properties of the LP Subproblems--A Summary

When y is fixed in problem (6.1) the result is an LP subproblem. In view of Theorem 4.3 we are interested in the continuity properties of LP problems. Probably the most general LP continuity result for the bounded case is expressed in the next proposition. Let y be fixed and define

$$\overline{Z}_y \equiv \{(b,d) \ \varepsilon \ E^m \ | \ F_y(\overline{G},\overline{H},b,d) \neq \emptyset\} \ . \tag{6.4}$$

The set $\overline{Z}_y$ is precisely the set of values of (b,d) for which

$$-\infty < \overline{v}_y(b,d) < \infty \ . \tag{6.5}$$

Theorem 6.4:

(a) $\overline{Z}_y$ is convex.

(b) $\overline{v}_y(c')$ is piecewise linear, concave, and continuous on E^p.

(c) $\overline{v}_y(b,d)$ is convex and continuous on $\overline{Z}_y$.

(d) v_y is continuous with respect to (c',G_1,H_1,b,d) at $\overline{\alpha}$ if $(\overline{b},\overline{d}) \ \varepsilon$ $\text{int}(\overline{Z}_y)$.

All statements of Theorem 6.4, except for piecewise linearity of $\overline{v}_y(c)$, follow as special cases of Theorems 5.1 and 5.3 which hold for convex programs. To see that $\overline{v}_y(c')$ is piecewise linear, let B denote the set of basic feasible solutions, x. Then

$$\bar{v}_y(c') = \underset{x \,\varepsilon\, B}{\text{minimum}} \{c'x\}$$

which is the minimum over a finite collection of linear functions. Thus, $\bar{v}_y(c')$ is a piecewise linear (and concave) function of c' with a finite number of "pieces." (By similarly looking at the dual problem it follows that $\bar{v}_y(b,d)$ is piecewise linear (and convex).)

These properties are applicable also to the unbounded case (that is, when X is not necessarily bounded) by redefining $\bar{Z}_y$ to be the set of (b,d) for which (6.5) holds, and by restricting c to the subset of E^n for which $-\infty < \bar{v}(c) < \infty$. These sets are also easily shown to be convex.

Statement (d) says that if any "small" perturbation in $(\bar{b},\bar{d})$ is feasible (i.e., admits a feasible solution at G,H), then a "small" perturbation in all of the parameters will induce only a small change in v_y. As a special case of (d) we have that $\bar{v}_y(G_1,H_1)$ is continuous at $(\bar{G}_1,\bar{H}_1)$ if $(\bar{b},\bar{d}) \,\varepsilon\, \text{int}(\bar{Z}_y)$; this is particularly intriguing in that continuity with respect to the elements of G_1 and H_1 requires a condition on (b,d) but none on (G,H). Indeed, the pattern established by (b) and (c) could lead one to an incorrect conclusion regarding continuity with respect to G and H -- although $\bar{v}_y(b,d)$ is continuous on a domain for (b,d) and $\bar{v}_y(c)$ is continuous on a domain for c, it is not true that $\bar{v}(G,H)$ is continuous on an analogous domain for (G,H). Consider the problem (taken from Gale [10]):

$$\begin{array}{ll} \text{maximize } x & \\ x \,\varepsilon\, [0,1] & \quad\quad (6.6) \\ \text{subject to: } ax \leq 0 & \end{array}$$

where "a" is a scalar. Any value for "a" from the domain $(-\infty,\infty)$ admits a finite optimal solution, yet the infimal value function is not

continuous at a = 0:

$$v(a) = \begin{cases} 1, & a \le 0 \\ 0, & a > 0 \ . \end{cases}$$

This example does not satisfy the interiority condition of statement (d). Viewing the constraint as $ax \le b$, $b = 0 \notin int \{b | 0 \cdot x \le b\} = (0, \infty)$.

In LP terminology, Theorem 6.4(d) says that the optimal value of a bounded LP problem is continuous with respect to all parameters (at the given point) if the resource vector lies in the interior of its domain of feasibility (defined with the technology and cost coefficients fixed). The usual conditions for LP continuity are expressed in terms of "regularity" or "boundedness of the primal and dual optimal solution sets" (see, for instance, Gale [10], Bereanu [3], Hoffman and Karp [15], and Williams [21]). If we let $W(\bar{\alpha})$ denote the optimal dual solution set at $\bar{\alpha}$, then $M(\bar{\alpha})$ and $W(\bar{\alpha})$ bounded imply that v for the LP is continuous at $\bar{\alpha}$ (see [15] for a proof). The condition given in 6.4(d) is simpler in the sense that it is (usually) easier to perceive and to apply when dealing with entire classes of problems. Furthermore, a similar condition (actually the same) is needed to assure continuity in the MIP context. We must emphasize, however, that this result relies on boundedness of X.

If X is not bounded, the interiority condition of 6.4(d) is not sufficient for continuity of the optimal value (with respect to technology coefficients). The following problem is taken from Bereanu [3]:

$$\begin{array}{c} \underset{x_1 \ge 0, x_2 \ge 0}{\text{minimize}} \quad x_1 \\ \\ \text{subject to:} \quad x_1 + ax_2 \ge b \ . \end{array} \tag{6.7}$$

Here, $v(a,b)$ is not continuous at $(0,1)$ since

$$v(a,1) = \begin{cases} 1, & a \le 0 \\ 0, & a > 0 \ . \end{cases}$$

Yet, at $\bar{a} \equiv 0$, $\bar{b} \equiv 1 \ \varepsilon \ \text{int}(\bar{Z}) = (-\infty,\infty)$. Hence, interiority is satis-fied. In this case, discontinuity is due to unboundedness of X, for if $a > 0$ and $b = 1$, an optimal solution is $(x_1^*,x_2^*) = (0,1/a)$ so that $x_2^* \to \infty$ as $a \to 0$.[*]/

6.3 Conditions for Continuity

A primary continuity property of linear MIP problems follows as a special case of Theorem 5.1(c) which applies to convex MIP problems.

Theorem 6.5: v is ε-continuous at $\bar{\alpha} \equiv (\bar{c},\bar{G},\bar{H},\bar{b},\bar{d})$ if $(\bar{b},\bar{d}) \ \varepsilon \ \text{int}(\bar{Z}_y)$ for some y that is ε-optimal at $\bar{\alpha}$.

This result states that if small perturbations in $(\bar{b},\bar{d})$ do not destroy feasibility when $\bar{y}$ (optimal at $\bar{\alpha}$) is fixed, then v is continu-ous at $\bar{\alpha}$. The computational usefulness of this result for continuity analysis is that one needs to address only the right-hand side param-eters (b,d).

The interiority condition in Theorem 6.5 is sufficient but not necessary for continuity of v (whereas it is both necessary and suffi-cient for LP problems). A simple example which shows that interiority is not necessary is the following:

---

[*]/Notice that, in fact, v is not even l.s.c. at $(0,1)$. Thus (6.7) demon-strates that Theorems 3.2, 3.4, 5.7, and 6.1, to name a few, are not valid without some sort of boundedness assumption. Hogan [16] shows that, with reference to the point $\bar{\alpha}$, it is sufficient to assume that there exists a neighborhood $N(\bar{\alpha})$ such that $\text{cl}[\ \underset{\alpha \varepsilon N(\bar{\alpha})}{\cup}\ F(\alpha)]$ is compact. This, of course, is weaker than the assumption that X is compact.

$$\text{minimize } 0$$
$$y = 0 \text{ or } 1$$

$$\text{subject to} \quad \alpha - 1 \le y \le \alpha .$$

Take $\bar{\alpha} = 1$. Thus, the constraint is: $0 \le y \le 1$. Both $y = 0$ and $y = 1$ are feasible and optimal, and

$$\bar{Z}_0 = \{\alpha | \alpha - 1 \le 0 \le \alpha\} = [0,1]$$
$$\bar{Z}_1 = \{\alpha | \alpha - 1 \le 1 \le \alpha\} = [1,2] .$$

Hence, $\bar{\alpha} = 1$ is not in the interior of $\bar{Z}_y$ for any y that is optimal at $\bar{\alpha}$, yet $v(\alpha)$ is obviously continuous at $\alpha = 1$.

A situation where the interiority condition of Theorem 6.5 is both necessary and sufficient is when $\bar{y}$, optimal at $\bar{\alpha}$, is unique.

Theorem 6.6: If $\bar{y}$ is uniquely optimal at $\bar{\alpha}$, then v is continuous at $\bar{\alpha}$ if and only if $(\bar{b}, \bar{d}) \in \text{int}(\bar{Z}_{\bar{y}})$.

Proof: Sufficiency is the content of Theorem 6.5. We will establish the contrapositive of the converse. For ease of exposition, let $\beta = (b,d)$. Suppose $\bar{\beta} \in \text{bd}(\bar{Z}_{\bar{y}})$ where $\bar{y}$ is optimal and unique at $\bar{\alpha}$. By (6.5) we see that $\bar{v}(\beta)$ is not continuous at $\bar{\beta}$ since $\bar{v}_{\bar{y}}(\bar{\beta}) < \infty$ for $\bar{\beta} \in \bar{Z}_{\bar{y}}$ and $\bar{v}_{\bar{y}}(\beta) = \infty$ for $\bar{\beta} \notin \bar{Z}_{\bar{y}}$. Therefore, there exists a sequence $\{\beta^k\}$, $\beta^k \to \bar{\beta}$, such that

$$\lim_{k \to \infty} \bar{v}_{\bar{y}}(\beta^k) \ne \bar{v}_{\bar{y}}(\bar{\beta}) . \tag{6.8}$$

Suppose y^k is optimal at β^k, i.e., at $\alpha^k = (\bar{c}, \bar{G}, \bar{H}, \beta^k)$. Since Y is a finite set, there must be a y^0 such that $y^k = y^0$ for infinitely-many k. That is, there exists a subsequence $\{\beta^{k_j}\}$ of $\{\beta^k\}$ such that y^0 is optimal on $\{\beta^{k_j}\}$. It follows from Theorem 6.2 that $\bar{Z}_{y^0}$ is closed so that $\bar{\beta} \in \bar{Z}_{y^0}$ since $\beta^{k_j} \to \bar{\beta}$ (i.e., y^0 is feasible at $\bar{\beta}$). Now, by Theorem

6.4(b), $\bar{v}_{y^0}(\beta)$ is continuous on $\bar{Z}_{y^0}$ so that

$$\lim_{k_j \to \infty} \bar{v}_{y^0}(\beta^{k_j}) = \bar{v}_{y^0}(\bar{\beta}) \ . \tag{6.9}$$

Since y^0 is optimal on $\{\beta^{k_j}\}$, $v_{y^0}(\beta^{k_j}) = v(\alpha^{k_j})$. (Recall that $\alpha^{k_j} \equiv (\bar{c},\bar{G},\bar{H},\beta^{k_j})$; hence, $\alpha^{k_j} \to \bar{\alpha}$.) Thus, (6.9) implies that

$$\lim_{k_j \to \infty} v(\alpha^{k_j}) = \bar{v}_{y^0}(\bar{\beta}) \ . \tag{6.10}$$

If v is continuous at $\bar{\alpha}$, it follows from (6.10) that

$$\bar{v}_{y^0}(\bar{\beta}) = v(\bar{\alpha}) \ . \tag{6.11}$$

But $\bar{v}_{y^0}(\bar{\beta}) = v_{y^0}(\bar{\alpha})$ and $v(\bar{\alpha}) = v_{\bar{y}}(\bar{\alpha})$ so that (6.11) implies $v_{y^0}(\bar{\alpha}) = v_{\bar{y}}(\bar{\alpha})$. This means that y^0 is optimal at $\bar{\alpha}$ since $\bar{y}$ is optimal, thus contradicting the assumption that $\bar{y}$ is unique. We conclude that v cannot be continuous at $\bar{\alpha}$. $\quad ||$

With regard to inequality constraints, the interiority condition could be replaced by strong feasibility, a condition that is relatively easy to observe. In the case of equality constraints, on the other hand, the concept of strong feasibility is not as useful or applicable. Nevertheless, for certain types of equality constraints it may make sense to speak of strong feasibility. In numerous applications, equality constraints involve only the discrete variables. That is, they often have the form $H_2 y = d$. Furthermore, for this form, H_2 and d usually are integral. An example is the set of configuration constraints: $\sum_{j \in J_i} y_j = I_i$, where I_i is an integer and $i = 1,2,\ldots,m_2$. The question of continuity does not apply to such parameters, for we are not interested in (small) perturbations of integer-valued parameters.

Consideration of such equality constraints is encompassed in the following statement.

Theorem 6.7: v is ε-continuous with respect to (c,G,b) at $(\bar{c},\bar{G},\bar{b})$ and $M_\varepsilon(\cdot)$ is closed with respect to (c,G,b) at $(\bar{c},\bar{G},\bar{b})$ if for some $\bar{y}$ ε-optimal at $\bar{\alpha} \equiv (\bar{c},\bar{G},\bar{H},\bar{b},\bar{d})$ there exists an $x \in U$ such that $\bar{G}_1 x + \bar{G}_2 \bar{y} < \bar{b}$.

Proof: (Equality constraints are encompassed by the set U.) The result follows directly from Theorem 4.5 since, by definition, $\bar{y}$ is s.f. at $\bar{\alpha}$ if $\bar{G}_1 x + \bar{G}_2 \bar{y} < b$ for some $x \in U$ (see paragraph 4.3). ||

Sometimes the constraints that appear in well-known models as equalities could be expressed as inequalities for the purpose of a continuity analysis. For instance, the demand constraints in facility location models are usually expressed as equalities because it is obvious that equality will hold in an optimal solution even if the constraints are treated as inequalities (demand must be met or exceeded).

Corollary 6.7.1: In the pure integer problem: $\min_{y \in Y} cy \mid Gy \leq b$, $v(\cdot)$ is continuous at $\bar{\alpha} \equiv (\bar{c},\bar{G},\bar{b})$ if and only if $\overline{Gy} < \bar{b}$ for some $\bar{y}$ optimal at $\bar{\alpha}$.

Proof: Necessity is obvious since if equality holds in any constraint, a change in $\bar{b}$ could render $\bar{y}$ infeasible (of course, directional continuity is possible). Sufficiency follows from Theorem 6.7.

6.4 Application Examples

The first example demonstrates the application of Theorems 6.5 and 6.6.

Example 6.1: Organizational Budgeting

99

$$\text{maximize} \quad \sum_{j=1}^{q} c_j y_j - c_0 \left(\sum_{i=1}^{p} x_i \right)$$
$$x_i \geq 0$$
$$y_j = 0 \text{ or } 1$$

$$\text{subject to:} \quad \sum_{j=1}^{q} a_{ij} y_j \leq L_i + x_i, \quad i = 1,2,\ldots,p \qquad (6.12)$$

$$\sum_{i=1}^{p} x_i \leq b \ .$$

A company with p divisions is considering undertaking certain of q different projects. Each project requires mixed participation of the different divisions. Each division has its own budget (in terms of money, personnel, etc.) and, in addition, the company has recognized a budget b to be allocated among the p divisions. The problem is to select a subset of the q projects and to allocate the budget b to the p divisions in a way that maximizes return.

The parameters may be interpreted as follows:

c_j = return (expected) on project j, (j = 1,2,...,q)

c_0 = per unit penalty for using company support (b)

a_{ij} = division i expense for its role in project j

L_i = division i budgetary capacity

b = available company support.

The variable y_j is 1 if project j is to be undertaken and 0 if not; the variable x_i is the amount of company support allocated to division i.

Suppose that y is held fixed at $\overline{y}$. Then x must satisfy $x \geq 0$, $a_i \overline{y} - L_i \leq x_i$ (i = 1,...,p), and $\sum x_i \leq b$, where $a_i \equiv (a_{i1},\ldots,a_{iq})$. The parameters L_i and b must satisfy

$$b \geq \sum_{i=1}^{p} \max \{0, a_i \overline{y} - L_i\} \qquad (6.13)$$

and $L_i \geq 0$ $(i = 1,\ldots,p)$. This means that

$$\overline{Z}_y = \{(b,L_1,\ldots,L_p) | L \geq 0, \ (6.13) \text{ holds}\} . \qquad (6.14)$$

By Theorem 6.5, the maximal value function v is continuous at
$(c_0,\ldots,c_q; \ a_{11},\ldots,a_{1q}; \ \ldots; \ a_{p1},\ldots,a_{pq}; \ L_1,\ldots,L_p; \ b)$ if $(b,L_1,\ldots,L_p)$
satisfies

$$b > \sum_{i \in I} (a_i\overline{y} - L_i)$$

$$(6.15)$$

$$(b,L_1,\ldots,L_p) > 0$$

where $I \equiv \{i | a_i\overline{y} - L_i \geq 0, \ i = 1,2,\ldots,p\}$. It can be assumed without loss
of generality that $L > 0$. Condition (6.15) means that with y fixed at
$\overline{y}$ there exists a feasible value of x such that $\Sigma x_i < b$.

If $b = 0$, problem (6.12) is a pure integer problem. In this case,
v is continuous with respect to all parameters (b excluded) if
$(L_1,\ldots,L_p) > 0$ and $L_i > a_i y$, $i = 1,\ldots,p$. That is, continuity holds
if each of the divisions requires less than its budgetary capacity (this
result is obvious).

In the general case ($b > 0$), continuity of v holds if not all of
the company support b is allocated in an optimal solution. If this con-
dition does not hold at a derived optimal solution, v might still be
continuous if $\overline{y}$ is not unique. Precisely, v is continuous at α if
(6.15) holds for some $\overline{y}$ that is optimal at α; if $\overline{y}$ is unique, v is con-
tinuous at α if and only if (6.15) holds.

The conclusions in the foregoing example are derived via an expli-
cit characterization of $\overline{Z}_y$. In some applications, deriving such a
characterization may prove to be an overly tedious exercise. It be-
hooves us to establish implicit means for determining interiority.

101

For the case where there are no equality constraints ($m_2 = 0$), the existence of an $x \in X$ for which $\overline{G}_1 x + \overline{G}_2 \overline{y} < \overline{b}$ implies that $\overline{b} \in \text{int}(\overline{Z-}_y)$. This strict inequality (in each component) is the condition that $\overline{y}$ is strongly feasible (s.f.) at $\overline{\alpha}$; (in the pure integer case, strong feasibility means that strict inequality holds in all constraints at $\overline{y}$). In this case, Theorem 6.5 applies: v is ε-continuous at $\overline{\alpha}$ if there is a $\overline{y}$ ε-optimal and s.f. at $\overline{\alpha}$.

Example 6.2:   Multicommodity Distribution System Design

The following distribution system design problem is addressed by Geoffrion and Graves [12].

$$
\begin{array}{l}
\underset{\substack{x \geq 0 \\ y,z=0,1}}{\text{Minimize}} \quad \sum_{ijk\ell \in N} c_{ijk\ell} x_{ijk\ell} + \sum_k [c'_k z_k + \sum_\ell c''_{k\ell} y_{k\ell}]
\end{array}
$$

subject to: (1) $\displaystyle\sum_{k\ell} x_{ijk\ell} \leq S_{ij}$, all $ij \ni S_{ij} > 0$

(2) $\displaystyle\sum_j x_{ijk\ell} = D_{i\ell} y_{k\ell}$, all $ik\ell \ni D_{i\ell} > 0$ (6.16)

(3) $\displaystyle\sum_k y_{k\ell} = 1$, all ℓ

(4) $L_k z_k \leq \displaystyle\sum_{i\ell} D_{i\ell} y_{k\ell} \leq U_k z_k$, all k

(5) Linear configuration constraints on y and z

where $N \equiv \{ijk\ell \mid S_{ij} > 0 \text{ and } D_{i\ell} > 0\}$

and $x_{ijk\ell} \equiv$ amount of commodity i shipped from plant j through

distribution center k to customer zone ℓ

$$
y_{k\ell} \equiv \begin{cases} 1, & \text{if center k serves zone } \ell \\ 0, & \text{otherwise} \end{cases}
$$

$$
z_k \equiv \begin{cases} 1, & \text{if a center is acquired at site k} \\ 0, & \text{otherwise.} \end{cases}
$$

Parameter definitions and a discussion of the model are given in [12].
(The problem is bounded since $x_{ijk\ell} \leq D_{i\ell} < \infty$ for all $i\ell$.) Notice that
(2) may be written as inequalities:

$$(2)' \qquad \sum_j x_{ijk\ell} \geq D_{i\ell}y_{k\ell}$$

if we assume that $c_{ijk\ell} > 0$, for then no $x_{ijk\ell}$ would be unnecessarily
large. Secondly, (3) and (5) do not involve (continuous) data to be
analyzed. We need, therefore, to address only constraints (1), (2)',
and (4), all of which are inequalities. The parameters of interest are
$S = [S_{ij}]$, $D = [D_{i\ell}]$, the vectors L and U, and the objective function
coefficients c, c', and c". Let $\bar{\alpha}$ denote the given values of these
parameters.

Assume that the total capacity for producing commodity i exceeds
the total demand for i. That is,

$$\sum_\ell D_{i\ell} < \sum_j S_{ij}, \quad \text{all} \quad i \ . \tag{6.17}$$

Let $(\bar{x},\bar{y},\bar{z})$ denote an optimal solution. For each i there is a j_i such
that $\sum_{k\ell} x_{ij_ik\ell} < S_{ij_i}$, for if not,

$$\sum_j (\sum_{k\ell} \bar{x}_{ijk\ell}) = \sum_j S_{ij} = \sum_{k\ell} (\sum_j \bar{x}_{ijk\ell})$$

$$= \sum_k D_{i\ell}\bar{y}_k, \quad \text{by (2)}$$

$$= \sum_\ell D_{i\ell}, \quad \text{by (3)}$$

which contradicts (6.17). For each i, let $\Delta_i \equiv S_{ij_i} - \sum_{k\ell} \bar{x}_{ij_ik\ell} > 0$.
Let n be a very large number, and define

103

$$x_{ijk\ell} = \overline{x}_{ijk\ell}(1 - \frac{1}{n}\Delta_i), \quad j \neq j_i$$

$$x_{ij_ik\ell} = \overline{x}_{ij_ik\ell} + (\frac{1}{n+1})\Delta_i . \tag{6.18}$$

It is clear that $x \geq 0$ and for n sufficiently large,

$$\sum_{k\ell} x_{ijk\ell} < S_{ij}$$

$$\sum_{j} x_{ijk\ell} > D_{i\ell}\overline{y}_{k\ell} . \tag{6.19}$$

We have established strong feasibility with respect to constraints (1)
and (2)'. We cannot establish the same for constraints (4) because
$z_k = 0$ forces equality (for all $x \in X$). However, the condition

$$L_k < \sum_{i\ell} D_{i\ell}\overline{y}_{k\ell} < U_k, \quad \text{all} \quad k \ni \overline{z}_k = 1 \tag{6.20}$$

together with (6.19) would permit the application of Theorem 6.7 since
for $\overline{z}_k = 0$, L_k, U_k, and D can be perturbed in any fashion without render-
ing $(\overline{y},\overline{z})$ infeasible; i.e., apply Theorem 6.5 to the constraints in (4)
for which $\overline{z}_k = 0$, and apply Theorem 6.7 to all other constraints. Alter-
natively, (6.19), (6.20), and the subsequent observation regarding those
k for which $\overline{z}_k = 0$, collectively imply that any small perturbations in
S, D, L, and U do not render $(\overline{y},\overline{z})$ infeasible. Continuity of v then
follows directly from Theorem 6.5.

The following conclusions regarding problem (6.16) are at hand
(ϵ-optimal $\overline{y}$ and $\overline{z}$ are assumed to be available).

(1) The optimal value function v is ϵ-continuous at $\overline{\alpha}$ with respect to
all parameters if

$$\sum_{\ell} D_{i\ell} < \sum_{j} S_{ij}, \quad \text{all} \quad i$$

$$L_k < \sum_{i\ell} D_{i\ell}\overline{y}_{k\ell} < U_k, \quad \text{all} \quad k \ni \overline{z}_k = 1 .$$

104

(2) If supply exceeds demand (for each product i) and

$$L_k = \sum_{i\ell} D_{i\ell} \overline{y}_{k\ell}, \qquad k \in K_1$$

$$\sum_{i\ell} D_{i\ell} \overline{y}_{k\ell} = U_k, \qquad k \in K_2 ,$$

then v is ε-continuous with respect to c, c', c", S, L_k for k $\notin$ K_1, U_k for k $\notin$ K_2, and with respect to decreases in L_k (k $\in$ K_1) and increases in U_k (k $\in$ K_2). If K_1 (K_2) is empty, then v is also ε-continuous with respect to decreases (increases) in D.

In this particular example the LP subproblem that results when the integer variables are fixed involves only a subset of the constraints, namely (1) and (2), and therefore the LP problem does not involve all of the parameters (namely, L and U). Obviously, continuity of the LP subproblems is not sufficient (nor is it necessary) for continuity in the MIP problem. For instance, in Theorem 4.3, the interiority condition (a) complements the condition that the LP subproblem (at an optimal $\overline{y}$) be continuous; the condition (a) is used to assure that perturbations in the parameters do not render $\overline{y}$ infeasible. Theorem 6.5 states that both of these conditions are satisfied by a single interiority condition in the linear case.

Our reason for not employing "regularity" or "primal and dual solution set boundedness" conditions should be clear. Although such conditions are sufficient for LP continuity, their imposition in the context of LP subproblems must be supplemented by another condition such as $\alpha \in int(A_y)$. In the MIP context, boundedness of the dual solution set is not a meaningful concept (see [1], [13], and particularly the conclusion in [14], regarding "reimputed dual prices").

105

CHAPTER 7

SOME NOTES ON MODEL FORMULATION

Conditions have been established in previous chapters that assure
that the optimal value of a bounded MIP problem varies continuously
with respect to the problem data. The practical usefulness of this
theory was demonstrated via application to several prominant classes of
problems. In particular, general models were analyzed for
capacitated facility location (4.9) and (4.13), dynamic lot sizing
(4.18), organizational or hierarchical capital budgeting (6.12), and
multicommodity distribution system design (6.16).

Knowledge that the optimal value v is continuous is assurance that
small perturbations in the data will create only a small change in the
optimal value. Knowledge that v is discontinuous at a point is a warn-
ing to the analyst: the model may be an inadequate representation of
the real problem; or the discontinuity may in fact be realistic, in
which case the analyst should like to know that particular changes in
data are responsible for the discrete behavior of v.

Given a model that is likely to exhibit discontinuous behavior
(based on continuity theory) or that has already been demonstrated to
be discontinuous (e.g., by solving, observing the critical data that
violate conditions for continuity, and resolving with the critical data
perturbed), the analyst may want to alter the problem in a way that
precludes discontinuity. This might be accomplished by simply revising
the data or by improving upon the model itself.

It was pointed out earlier that it is considerably more difficult
to formulate a realistic MIP model than an LP model; the reason stems

from the fact that LP models possess more favorable continuity proper-
ties. Therefore, MIP problems are particularly indigent of efficient
methods for formulating/reformulating realistic models. Naturally, the
development of such modeling techniques must acknowledge the theory of
continuity.

One way to add realism to a model, thereby preventing undesirable
discontinuities, is to treat certain critical parameters as explicit
variables. A technique suggested by Williams [22] is to incorporate
complicating constraints into the objective function via penalty func-
tions. We will see shortly that these two ideas can be viewed as equiv-
alent. In light of Theorems 3.12, 5.1, and 6.5, continuity/discontinui-
ty of v can be assessed by treatment of only the right-side data. These
data, then, are the primary candidates for being represented by explicit
variables. Replacement of data by variables will be referred to as
relaxation of data (paragraph 7.1). Removal of constraints via penal-
ties is included also under this title.

Another technique, when applicable, adds realism to the model and
in so doing, improves the inherent continuity properties. This is the
treatment of discrete variables as quasi-integer (see Armstrong and Sinha
[2]). A variable y_i is quasi-binary if it must satisfy the condition

$$y_i = 0 \quad \text{or} \quad \ell_i \leq y_i \leq u_i \tag{7.1}$$

where $\ell_i \leq 1 \leq u_i$. Use of quasi-binary variables in place of 0-1 variables
could, in many applications, be more realistic; the improvement in con-
tinuity behavior is discussed in paragraph 7.2. Replacing an integral
variable value (e.g., 1) by an interval will be referred to as relaxa-
tion of integrality.

107

7.1 Relaxation of Right-Side Data

Constraint formulations that are adequate in an LP model may not be acceptable in an MIP model in that unrealistic discontinuities can occur. Consequently, greater care must be taken in formulating an MIP problem in order to achieve an acceptable representation of the real problem.

Typically, adding realism to a model results in improved continuity properties, and vice versa. Perhaps the most obvious means for improving a model is replacement of a data point by a (usually bounded) variable. In a capital investment model, the budgetary constraint appears as

$$\Sigma \ a_i y_i \leq b \ .$$

It was illustrated in Chapter 1 that such a constraint permits discontinuities in an optimal solution value; in effect, an infinite penalty is assigned to all investment plans that exceed the budget b by even the slightest amount. It seems more realistic to allow b to vary and to represent the cost of any deviation from b in the objective function.

To exemplify these observations, we will show how a discontinuity in a capital investment problem can be alleviated by replacing an inadequate budgetary constraint by one which is more realistic. Consider the following problem with four independent investment opportunities:

$$\text{maximize } [200y_1 + 160y_2 + 60y_3 + 40y_4]$$
$$\text{subject to: } 41y_1 + 39y_2 + 20y_3 + 10y_4 \leq b \qquad (7.2)$$
$$y_i = 0 \text{ or } 1; \quad i = 1,2,3,4 \ .$$

108

For b = 80 the optimal solution is $y^* = (1,1,0,0)$ for an optimal return on investment of $v = 360$. But if $b = 79.9$, y^* is infeasible; the optimal return is 300 with $y = (1,0,1,1)$. Hence, v is discontinuous as a function of b at b = 80. [In fact, v is discontinuous at exactly 12 points: b = 10, 20, 30, 39, 41, 51, 61, 71, 80, 90, 100, and 110. On the other hand, it sounds more reassuring to say that v is continuous at all points b > 0 except for these 12.]

The discontinuity at b = 80 does not reflect the true nature of the investment problem. Surely it must be possible to stretch a budget of 79.9 to 80 if it means a return of 360 instead of 300. The small amount of additional capital might be borrowed in order to follow the superior plan. Figure 7.1 depicts a simple representation of the cost (value) of additional (excess) capital. From the figure,

$$c = \begin{cases} r_1(b - \overline{b}), & \text{if } b \geq \overline{b} \\ r_2(\overline{b} - b), & \text{if } \overline{b} > b . \end{cases} \tag{7.3}$$

c (cost of additional capital)
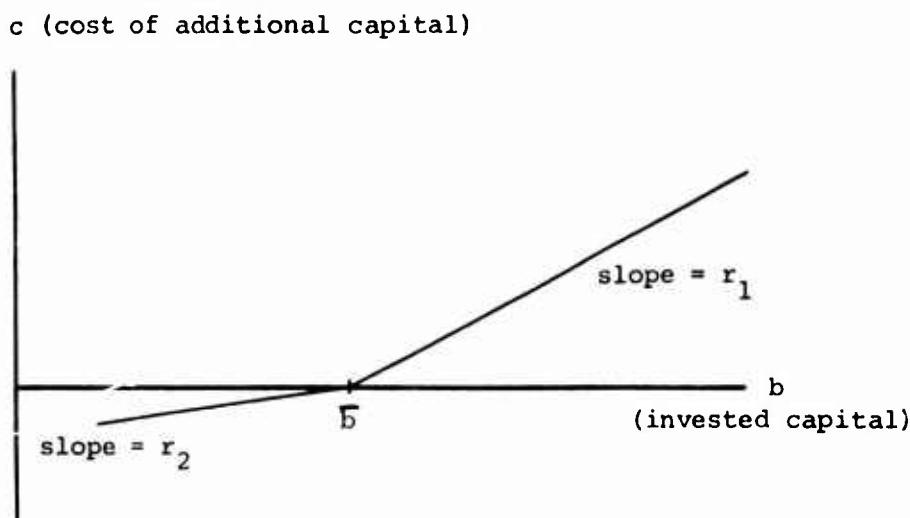
Figure 7.1. Example of Additional Capital Cost

We introduce the continuous variables:

$$z_1 \equiv \begin{cases} b - \overline{b}, & \text{if } b > \overline{b} \\ 0, & \text{if } b \leq \overline{b} , \end{cases} \qquad z_2 \equiv \begin{cases} \overline{b} - b, & \text{if } \overline{b} > b \\ 0, & \text{if } \overline{b} \leq b . \end{cases} \qquad (7.4)$$

Thus, $b = \overline{b} + z_1 - z_2$. Incorporating (7.3) and (7.4) in the model (7.2), we obtain

$$\text{maximize } [200y_1 + 160y_2 + 60y_3 + 40y_4 - r_1z_1 + r_2z_2]$$

$$\text{subject to: } 41y_1 + 39y_2 + 20y_3 + 10y_4 - z_1 + z_2 \leq \overline{b}$$

$$z_1 \geq 0, \quad z_2 \geq 0 \qquad (7.5)$$

$$y_i = 0 \text{ or } 1; \quad i = 1,2,3,4 .$$

The optimal value function for this problem will be denoted by $v^+(b)$. (Equality holds in the budget constraint if $r_2 > 0$. Inequality is used to permit the case where $r_2 = 0$.)

It is immediately clear that the optimal value $v^+(b)$ for formulation (7.5) is continuous at any given point $\overline{b}$ -- recall that v^+ can be discontinuous only at points $\overline{b}$ where some $y \in Y$ becomes infeasible (i.e., y is feasible for $b \geq \overline{b}$ but infeasible for $b < \overline{b}$), but in formulation (7.5) every $y \in Y$ is feasible; hence, there can be no discontinuities. Normally, one would place upper bounds on the z variables. In this case, discontinuity can occur at a point $\overline{b}$ only if z_1 (or z_2) is at its upper bound in every solution that is optimal at $\overline{b}$ (this follows from Theorem 6.7). The favorable continuity properties of (7.5) make the model more realistic and, depending on the provisions made in the code, could make the problem easier to solve than problem (7.2).

The graph of $v(b)$ for problem (7.2) is the discontinuous step function shown by the solid lines in Figure 7.2. This corresponds to $r_1 = \infty$
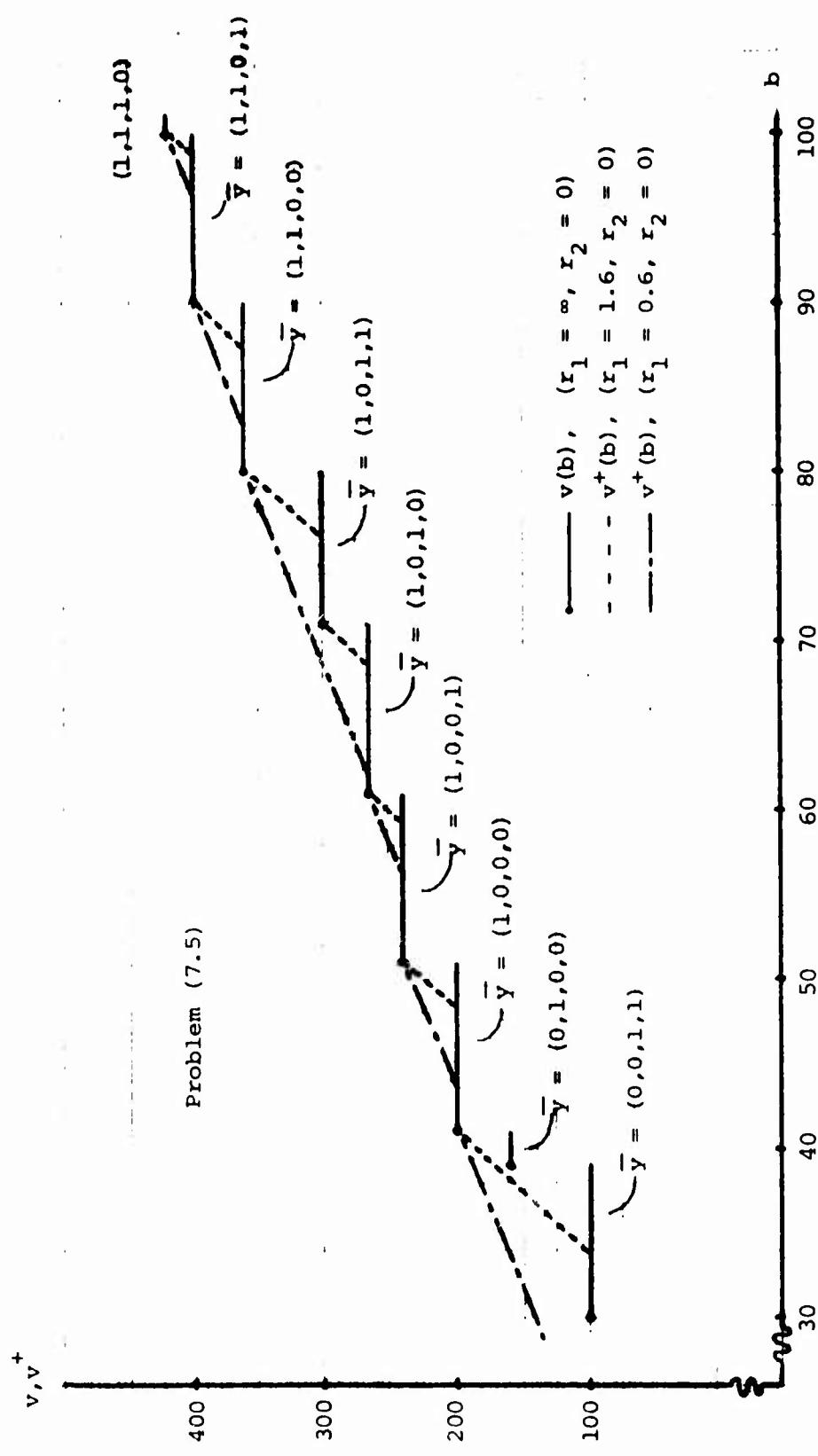
Figure 7.2. The Optimal Value Function for Problems (7.2) and (7.5).

111

and $r_2 = 0$ (see Figure 7.1) in problem (7.5). By comparison, Figure 7.2 shows also the optimal value function for problem (7.5) for the two cases, $r_1 = 1.6$ and $r_1 = 0.6$. In the latter cases, $v(b)$ is a continuous function. In the more general case where there are several constraints and b is a vector, a plot similar to Figure 7.2 is possible by permitting only directional changes in b.

The consideration of financing extensions to capital budgeting problems is not a new idea: a thorough development of such extensions has been presented by Bernhard [5]. A similar approach was taken by Davis and Larson [7] in the selection of multibenefit civil works projects.

The concept just illustrated can easily be extended to a more general case. Consider again problem (5.1). This problem is augmented as follows by introducing the continuous vector variables z and w.

$$
\begin{cases}
\text{Minimize } f(c;x,y) + p(z,w) \\
(x,y) \in U \\
\text{subject to: } G(a;x,y) - z \leq b \\
\qquad\qquad H(a;x,y) - w = d
\end{cases}
\tag{7.6}
$$

where U is a bounded set and p is an arbitrary bounded function of z and w; z is conformable with b, and w with d. It is further assumed that $p(0,0) = 0$. The function p can be viewed as a penalty for violating the constraints.

The optimal value function for problem (7.6) will be denoted by $v^+(\alpha)$ where $\alpha = (c,a,b,d)$. As before, A denotes the set of data points α for which (7.6) has a finite optimal value. The functions f, G, and H are assumed to be continuous on $A \times X \otimes Y$.

112

<u>Theorem 7.1</u>: $v^+(\cdot)$ is continuous on A.

<u>Proof</u>: Let $\overline{\alpha} \in A$ be arbitrary. Since $U \subseteq X \times Y$ is bounded, $\exists\ M < \infty \ni$ $\forall\ (x,y,z,w) \in F(\overline{\alpha})$, $\|z\| \le M$ and $\|w\| \le M$. Hence, the domain for (x,y,z,w) can be viewed as bounded. Suppose $(\overline{x},\overline{y},\overline{z},\overline{w})$ is optimal at $\overline{\alpha} \in A$. If (b,d) is perturbed to any value, say (b',d'), then $(\overline{x},\overline{y},z',w')$ is feasible at $\overline{\alpha}$ where $(z',w') \equiv (\overline{z},\overline{w}) + (b,d) - (b',d')$; continuity of v^+ at $\overline{\alpha}$ therefore follows from Theorem 5.1(c). $\quad\quad ||$

The addition of variables (z,w) can be viewed as replacing the right-side data (b,d) by variables $(\xi,\eta) \equiv (b,d) + (z,w)$. The variables (z,w) will be referred to as elastic data.

In practice it may be more meaningful to subject the elastic data to upper and lower bounds, especially when p is approximated by a simple form (e.g., linear or piecewise linear). In this case, $v^+(\cdot)$ can be discontinuous at a point if one of the elastic datum is at a bound in all optimal solutions. Now discussing the improved behavior of $v^+(\cdot)$ over that of $v(\cdot)$ is more difficult -- we need to address the notions of how many points of discontinuity there are and the "size" of a dis-continuity (i.e., the magnitude of a discrete jump in optimal value).

If an upper bound were placed on z in example problem (7.5) then the plots of v^+ in Figure 7.2 could possibly display some discontinui-ties; however, the number of discontinuities would tend to be fewer than for the step function $v(b)$ and the sizes of such discontinuities would be smaller as the bounds on z (i.e., on b) are made looser. For example, $v^+(b)$ with $r_1 = 0.6$ ($r_2 = 0 = z_2$) remains continuous for an upper bound of 5 on z_1 and is discontinuous only at b = 39 and 41 (with jumps of about 20) for an upper bound of 3.

The use of elastic data variables (z,w) to represent deviations from goals is characteristic of goal programming. As an example, federal funding of R & D projects usually involves several goals: maximize a measurable return, maximize coverage of various technological areas, expend the available annual budgets, maximize personnel involvements, etc. The selection/rejection of projects corresponds to integer 0-1 variables and funding levels correspond to continuous variables. This type of problem could be cast in the following simple goal programming format:

$$
\begin{cases}
\displaystyle \min_{z',z''\geq 0} \sum_{i=1}^{m} u_i(c_i' z_i' - c_i'' z_i'') \\[2ex]
\text{subject to:} \quad H_1 x + H_2 y + z'' - z'' = b \\[2ex]
\qquad\qquad (x,y) \in U
\end{cases}
\tag{7.7}
$$

where $U \subseteq X \times Y$, H_1 and H_2 are matrices conformable with the m-dimensional vector b, and u_i is a utility or priority associated with a deviation from the goal b_i. The set U accommodates side constraints.

The optimal value function v for problem (7.7) is continuous with respect to all data (u, c', c", H_1, H_2, b). This is observed immediately from Theorem 7.1. Consequently, as would be hoped, a goal programming formulation of a mixed integer problem assures that small modification of the goals b cannot result in an incommensurate change in the optimal solution value. Moreover, there is an obvious upper bound on the optimal value v' at a perturbed point b': writing b'-b as s'-s" where $s' \geq 0$ and $s'' \geq 0$, we have

$$
v' \leq v + \sum u_i(c_i' s_i' - c_i'' s_i'')
$$

where v is the optimal value at b (all other data fixed). Thus, the

114

maximum deviation of v' from v can be characterized in terms of the deviation of b' from b.

7.2 Relaxation of Integrality

Consider the following linear (mixed) quasi-binary problem:

$$
\begin{cases}
\text{maximize} \quad \sum_{i=1}^{n} c_i y_i \\
\text{subject to:} \quad \sum_{i=1}^{n} a_{ij} y_i \le b_j \qquad (j = 1,2,\ldots,m) \\
y_i = 0 \text{ or } \ell_i \le y_i \le u_i \quad (i = 1,2,\ldots,n)
\end{cases}
\tag{7.8}
$$

where $u \ge 1$ and $0 \le \ell \le 1$. This is a generalization of the pure integer problem where $\ell_i = u_i = 1$, $(i = 1,\ldots,n)$, which is typical in investment planning or R & D project selection.

Interpreted as a model of R & D project selection, the relaxed integrality constraints of (7.8) provide for greater realism in that projects can be accepted at variable levels. Consequently, projects with return functions $c_i(y_i)$ such as those depicted in Figures 7.3(a), (b), and (c) can be accommodated; the level of acceptance (amount of funding, etc.) of a project need not be fixed. This appears to be a much more realistic formulation of project (or portfolio) selection type problems and one can expect that it possesses much more favorable continuity properties than does the counterpart pure integer problem. There is also some evidence, which we will discuss shortly, that (7.8) is less difficult to solve than the pure integer problem ($\ell_i = u_i = 1$ for all i).

In the pure integer problem, $v(\alpha)$, $\alpha \equiv (c,a,b)$, is continuous at $\bar{\alpha}$ if and only if $\sum a_{ij} \bar{y}_i < b_j$, $(j = 1,\ldots,m)$ for some $\bar{y}$ that is optimal at $\bar{\alpha}$ (Corollary 6.7.1). It is typical in pure integer problems that an

115
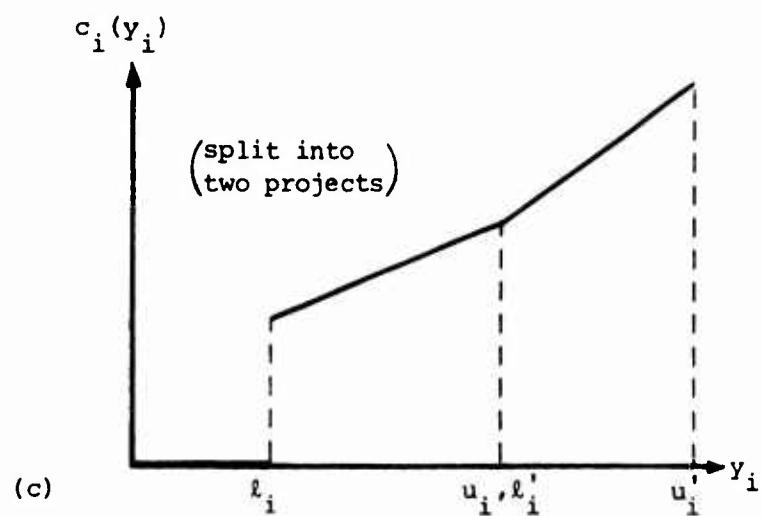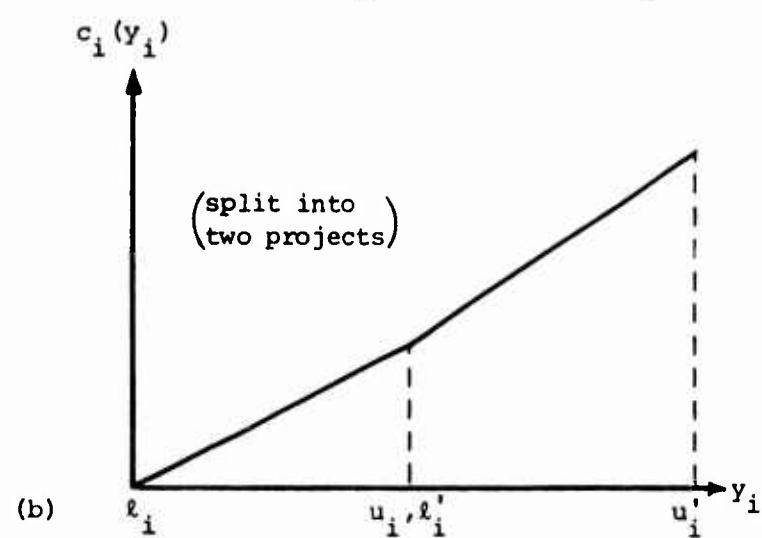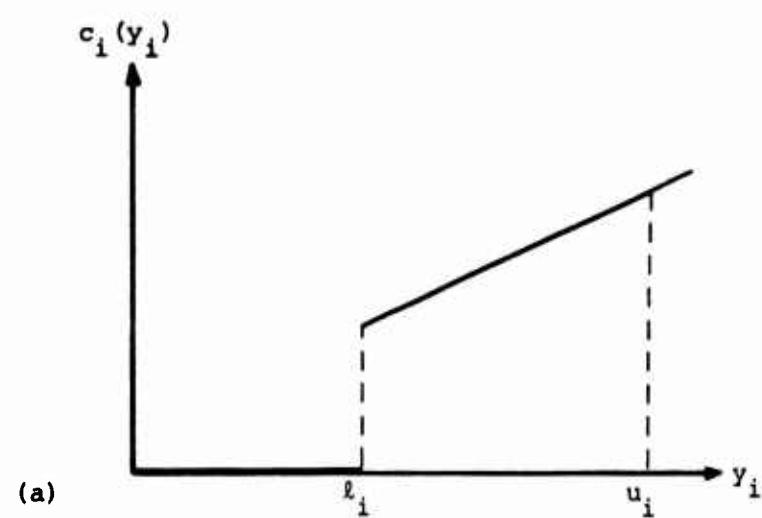
Figure 7.3. Example R & D Project Return Functions

116

optimal solution $\bar{y}$ is uniquely optimal (and even uniquely ϵ-optimal for reasonably small ϵ). In this case, v is discontinuous at all points for which equality holds in any one of the constraints (at the optimal solution).

In problem (7.8), however, (assuming $\ell \neq u$), v is discontinuous at a given data point only if (at every optimal solution) equality holds in some constraint j, for which $y_i = 0$ or ℓ_i (u_i) for all i such that $a_{ij} > 0$ ($a_{ij} < 0$). That is, v is discontinuous only if for some binding constraint an arbitrarily small reduction in the right-side would cause all optimal y to become infeasible (Theorem 6.7). Although v(b) generally is not continuous over all interesting values for b, points of discontinuity tend to be fewer in number and smaller in size than for the counterpart pure integer problem. This is illustrated in Figure 7.4 for example problem (7.2).

Both the relaxed problem and the pure integer problem are represented in Figure 7.4: the step function is the optimal value function for the pure integer problem (7.2), and the much smoother function -- which, of course, dominates the step function -- is the optimal value for the quasi-integer problem where y is relaxed by ±10%; i.e., $y_i \in \{0,1\}$ in (7.2) is replaced by $y_i = 0$ or $.9 \leq y_i \leq 1.1$. Only one discontinuity occurs for the quasi-integer problem (b $\approx$ 72); the discrete jump in v^+ is approximately 12 units, compared with a jump of 60 in the step function v (between the same solution points, i.e., b = 80). Again, for the case where there are several constraints and b is a vector, two-dimensional plots similar to Figure 7.4 are possible by considering directional changes in b; however, the main point of Figure 7.4 here is

117
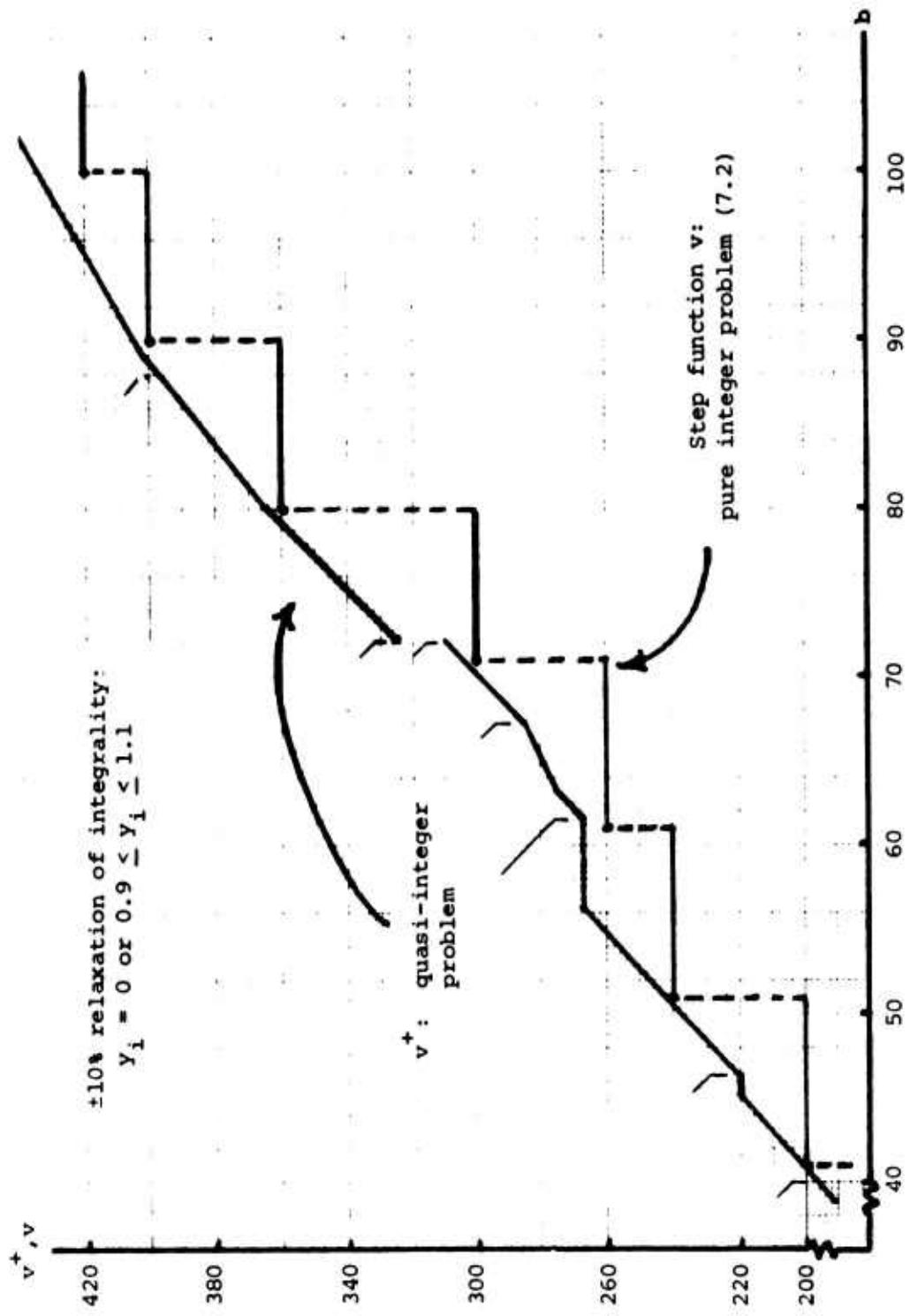
Figure 7.4. Relaxation of Integrality (±10%) in Problem (7.2)

to illustrate the improvement in continuity properties that can be achieved by allowing integer variables to be quasi-integer.

Armstrong and Sinha [2] describe modifications to a branch-and-bound algorithm of the Beale and Tomlin type for handling quasi-integer problems. They report that the modifications are minimal and no appreciable increase in storage is needed. The algorithm is applied to a menu-planning problem and computational results are given for ten problems: the quasi-integer problem ($\pm 10\%$) required on the average 20% less run time and over 40% fewer branches. Petersen's [18] investment problems were also studied: the quasi-integer problem (+10%) required 62% fewer branches in total for the seven problems. (This improvement is understandable: LP bounds are tighter in the quasi-integer case; one would expect branching to reduce as the allowable fluctuation in the variables is increased.)

7.3 Summary and Conclusions

Two general notions have been addressed. The first is that simple extensions to common types of models can lead to greatly improved continuity properties, and consequently, more realistic and better behaved models. Second, such added sophistication can sometimes actually make the problem easier to solve. This appears to be the case with relaxation of integrality (quasi-integer variables). When applicable, quasi-integer formulations can provide greater model fidelity and a reduced expense in obtaining solutions; sensitivity of v with respect to data is more meaningful and easier to assess, due to the "smoothness" of the optimal value.

Replacing data by variables (data relaxation) is a technique that

119

can add fidelity to the model; this results in improved continuity

properties. In particular, replacing right-side resource data by vari-

ables results in an optimal value function that is continuous with res-

pect to all data. This is comforting information in many applications

since such behavior is more realistic, and secondly, additional sensi-

tivity information is easier to obtain. The impact of this technique

on problem solution time has not been investigated.

BIBLIOGRAPHY - PART I

1. Alcaly, R.E., and A.K. Klevorick, "A Note on the Dual Prices of Integer Programs," _Econometrica_, Vol. 34, No. 1 (January 1966).

2. Armstrong, R.D., and P. Sinha, "Application of Quasi-Integer Programming to the Solution of Menu Planning Problems with Variable Portion Size," _Management Science_, Vol. 21, No. 4 (December 1974).

3. Bereanu, B., "The Continuity of the Optimum in Parametric Programming and Applications to Stochastic Programming," Report No. 7305, Center of Mathematical Statistics of the Academy of the Romanian Socialist Republic, Bucharest, July 1973.

4. Berge, C., _Topological Spaces_, Macmillan, New York, 1963.

5. Bernhard, R.H., "Mathematical Programming Models for Capital Budgeting - A Survey, Generalization, and Critique," _Journal of Financial and Quantitative Analysis_, Vol. 4 (June 1969).

6. Dantzig, G.B., J. Folkman, and N. Shapiro, "On the Continuity of the Minimum Set of a Continuous Function," _J. Math. Anal. and Appl._, Vol. 17 (1967).

7. Davis, R.E., and R.E. Larson, "Optimum Selection of Multibenefit Civil Works Projects," _Proceedings of the Sixth Annual Hawaii Conference on Systems Science_, January 1973.

8. Debreu, G., _Theory of Value_, Cowles Foundation Monograph No. 17, John Wiley, New York, 1959.

9. Evans, J.P., and F.J. Gould, "Stability in Nonlinear Programming," _Operations Research_, Vol. 18, No. 1 (January-February 1970).

10. Gale, D., _The Theory of Linear Economic Models_, McGraw-Hill, New York, 1960.

11. Geoffrion, A.M., "Elements of Large-Scale Mathematical Programming," _Management Science_, Vol. 16, No. 11 (July 1970).

12. _____, and G.W. Graves, "Multicommodity Distribution System Design by Benders Decomposition," _Management Science_, Vol. 20, No. 5 (January 1974).

13. Gomory, R.E., and W. Baumol, "Integer Programming and Pricing," _Econometrica_, Vol. 28, No. 3 (July 1960).

14. Hespel, P.P., "Reimputing the Dual Variables in Integer Programming," Discussion Paper No. 6707, Center for Operations Research and Econometrics, Universite Catholique de Louvain, Heverlee, Belgium, May 1967.

15. Hoffman, A.J., and R.M. Karp, "On Nonterminating Stochastic Games," _Management Science_, Vol. 12, No. 5 (January 1966), Appendix.

16. Hogan, W.W., Jr., "Point-to-Set Maps in Mathematical Programming," _SIAM Review_, Vol. 15, No. 3 (July 1973).

17. Meyer, R., "The Validity of a Family of Optimization Methods," _SIAM J. Control_, Vol. 8, No. 1 (February 1970).

18. Petersen, C.C., "Computational Experience with Variants of the Balas Algorithm Applied to the Selection of R&D Projects," _Management Science_, Vol. 13 (1967).

19. Rockafellar, R.T., _Convex Analysis_, Princeton University Press, Princeton, New Jersey, 1970.

20. Wagner, H.M., _Principles of Operations Research_, Prentice-Hall, Englewood Cliffs, New Jersey, 1969.

21. Williams, A.C., "Marginal Values in Linear Programming," _SIAM J. App. Math._, Vol. 11, No. 1 (March 1963).

22. _____, "Some Modeling Principles for M.I.P.'s," paper presented at the VIII International Symposium on Mathematical Programming, Stanford University, August 1973.

23. Zangwill, W.I., "Convergence Conditions for Nonlinear Programming Algorithms," _Management Science_, Vol. 16, No. 1 (September 1969).

24. _____, _Nonlinear Programming: A Unified Approach_, Prentice-Hall, Englewood Cliffs, New Jersey, 1969.

PART II

SENSITIVITY THEORY AND ANALYSIS

Chapters 8-11

- Introduction to Part II
- Sensitivity to Objective Function Data
- Right-Side Analysis: Basic Concepts and Properties
- Sensitivity to Right-Side Data: Analysis Methods

CHAPTER 8

INTRODUCTION TO PART II

The topic of continuity and continuity analysis studied in Part I addresses the qualitative concerns of sensitivity in MIP problems. We now turn to the more quantitative aspects. A primary objective is to discover data perturbations that take advantage of discontinuities, if any, to net a substantial improvement in the optimal value v. Of particular interest are techniques for bounding and approximating v.

8.1 General Considerations

The first consideration in applying and developing methods for sensitivity analysis is to identify the quantity and nature of information that is of practical use to the analyst. The derived information must be digestible and straightforward to interpret.

A major objective in studying sensitivity to data, especially nonparametric data, is to discover reasonable data perturbations that result in substantial improvements in optimal value. The sensitivity information, then, consists of a set of "alternative" data points, together with their respective improved optimal values and solutions (referred to as alternative solutions). One approach to obtaining this information is to permit the data to fluctuate from the nominal for a price. This requires that the analyst be able to "cost out" the economic (or social, environmental, etc.) consequences of perturbing the data.

It is necessary to assume that there exists a criterion for preferring one alternative solution over another. In some cases this may be difficult to quantify; however, it is most likely that the primary criterion is the dollar cost of additional resources required for the

124

alternative solution. An alternative solution that grants an improve-

ment in optimal value which exceeds the consequential costs of perturb-

ing the data is said to provide a net improvement.

Some flexibility in considering nonquantifiable preferences is

provided by generating several alternative solutions that yield net

improvements based on resource costs represented explicitly in the

model. The analyst may subjectively compare alternatives with regard

to costs that are too complex to be quantified.

Depending on the nature of the data, reasonably useful information

may be obtained by considering only directional changes in the data.

If p is a direction of interest in perturbing the objective function

data c, then one can address the parametric data $c+\theta p$ where θ is a

scalar and p and c are conformable. Thus, the optimal value v is a

function of the scalar variable θ and in principle, $v(\theta)$ could be plotted

as a two-dimensional graph. This functional relationship is obviously

much easier to approximate and interpret than the function $v(c)$ where c

is allowed to vary arbitrarily. In general, all data α could be param-

etrized on θ; the parametric relationship $\alpha(\theta)$ could be a nonlinear

(but continuous) function.

Sensitivity to data that might be perturbed in arbitrary fashion,

which will be referred to as "nonparametric data," cannot adequately be

represented parametrically in the sensitivity analysis. One might

imagine obtaining several plots of $v(\theta)$ corresponding to different

parametrizations, but such an attempt to obtain sufficient information

is costly and presents an unwieldy amount of material to digest. In

this case, a different approach must be taken.

Methods for deriving sensitivities to the several types of problem data -- objective function data (c), right-side data (b), and constraint function or left-side data (a) -- are necessarily different because the optimal value behaves differently. The optimal value v is continuous with respect to c but is generally discontinuous with respect to b and a.

Simultaneous fluctuation of all data is not given serious attention in this treatise because little can be gained in terms of sensitivity information from such an analysis. Allowing perturbations in both the left-side and the right-side data would produce alternative solutions that could also result from perturbations only in the right-side data. There appears to be little need to study effects of simultaneous changes in a and b. Studying the combined effects of perturbations in cost data c and right-side data b also buys little additional information because sensitivities to b and c to a large degree are incompatible: small changes in c can cause only small changes in v, whereas small changes in b can cause discrete jumps in v. Generally speaking, then, it is likely to be sufficient to limit sensitivity analyses to one type of data at a time -- right-side, left-side, or objective function data.

Analysis of left-side data is a far less attractive problem for study than right-side analysis, for several reasons. The left-side data, being a two-dimensional array, is comparatively bulky; its simultaneous and arbitrary perturbation would lead to an overwhelming amount of information that is neither practical to obtain nor to digest. If a particular perturbation of the left-side data admits an improved solution, then generally there is an infinite number of alternative perturbations that will admit the same improved solution. Moreover, there must

necessarily be a perturbation of right-side data alone that will admit
the same solution, suggesting that a search for alternative improving
solutions can be carried out by studying only the right-side data instead
of all constraint data. Let us illustrate this point. The constraint

$$41y_1 + 39y_2 + 20y_3 + 10y_4 \leq 78 ,$$

for example, precludes the binary solution $(1,1,0,0)$. This solution
would be feasible if $b \equiv 78$ were increased to 80. Alternatively, the
first two coefficients could be reduced respectively from 41 and 39 to
40 and 38; or to 39, 39; or to 40.5, 37.5; etc.

8.2 A Basic Approach

In devising a detailed approach to sensitivity analysis one natur-
ally seeks to obtain the desired kind and amount of information in the
computationally least expensive way. Two situations can exist: either
attractive alternative solutions exist, or they do not. In the latter
case, the fact that no reasonable data perturbations provide a substan-
tial improvement in $v(\cdot)$ hopefully can be discovered relatively easily,
thereby alleviating the need for any further analysis. In the former
case, the discovery of alternative solutions requires and merits a more
extensive computational effort.

We recommend approaching sensitivity analysis via the following
basic steps:

(1) Determine bounds on the improvement possible due to reason-
able data perturbations; stop if the potential improvement is
small.

(2) Attempt to derive alternative solutions that provide substan-
tial improvements in optimal value.

(3)   Analyze and compare the alternative solutions in order to

       arrive at a decision concerning which to implement or, possi-

       bly, concerning needed modifications to the model.

Steps (1) and (2) are addressed in detail in later chapters; step

(3), however, is itself a vast subject for study$^{*/}$ and is beyond the

scope of this treatise.

This approach applies primarily to the analysis of data that poten-

tially give rise to discontinuity in v (e.g., b). In cases where the

data perturbations of interest permit a continuous and tractable optimal

value function (e.g., parametric cost data), an approximation of v(·)

may be a reasonable objective.

8.3 Content and Summary

Sensitivity to objective function data c is studied in Chapter 9.

The optimal value is a concave function of the data c. In the linear

parametric case, v as a function of θ is piecewise-linear and concave

and therefore can be expressed as the pointwise minimum of finitely many

linear functions. The concavity property permits straightforward approx-

imation of v by quite simple bounding techniques. The content of Chap-

ter 9 is summarial in nature and consists mostly of evident corollaries

of the concavity property.

Methods for right-side analysis derive primarily from a few basic

concepts and techniques; these are presented in Chapter 10. Lower bounds

on v are provided from solutions to relaxed versions of the MIP problem;

upper bounds derive from feasibility and monotonicity properties.

Alternative methods of right-side analysis are studied in Chapter

---

$^{*/}$Comparison of alternative solutions could involve considerations
that are nonquantifiable and/or too complex for explicit expression in
the model.

11. Two methods, in particular, are emphasized: "rounding" of a non-integer solution obtained from a relaxation (of the domain) of the MIP, and relaxing the right-side data b by treating b as variable. A class of MIP problems is identified for which rounding of a solution to the continuous-variable relaxation, obtained by dropping integrality restrictions on the integer variables, always yields a solution that is optimal in the MIP at some perturbed right-side b'.

The second approach involves replacing b by b+z where z is variable and incurs a penalty $p(z)$. A solution to this problem produces an alternative solution to the MIP (i.e., a solution that is optimal at a perturbed right-side b'). A collection of alternative solutions can be generated by varying the penalty $p(\cdot)$.

8.4 Notation

The bounded mixed integer program in its most general form is given by

$$
\begin{aligned}
&\text{minimize } f(c;x,y) \\
&x\epsilon X, y\epsilon Y \\
&\text{subject to: } (x,y) \ \epsilon \ F(a)
\end{aligned}
\qquad (8.1)
$$

where X is compact, Y is a finite set, and F(a) is closed. The data c and a (or c, a, and b if right-side data are introduced) are identified with α; the optimal value function will be expressed as $v(\alpha)$. The notation $\overline{v}(c)$ will denote $v(c,\overline{a})$ where $\overline{a}$ is considered fixed; similarly, $\overline{v}(a) \equiv v(\overline{c},a)$ where $\overline{c}$ is held fixed.

For (linear) parametric objective function data we define

$$
\hat{v}(\theta) = \text{minimum } f(\overline{c}+\theta\overline{p};x,y) \,\big|\, (x,y) \ \epsilon \ F(\overline{a}) \qquad (8.2)
$$
$$
x\epsilon X, y\epsilon Y
$$

where $\overline{c}$, $\overline{p}$ and $\overline{a}$ are fixed. The same notation, $\hat{v}(\theta)$, will be used in

dealing with parametric right-side data and will therefore need to be read in context.

The notation $v_y(\cdot)$ will denote the optimal value of the continuous-variable subproblem that results when y is held fixed.

The ε-optimal solution set corresponding to $v(\alpha)$ is:

$$M_\varepsilon(\alpha) \equiv \{(x,y) \ \varepsilon \ X \times Y \mid (x,y) \text{ is } \varepsilon\text{-optimal at } \alpha\} . \qquad (8.3)$$

For $\varepsilon = 0$ we will simply write $M(\alpha)$. The set $\hat{M}(\theta)$ corresponds similarly to $\hat{v}(\theta)$, $\overline{M}(c)$ to $\overline{v}(c)$, etc.

The variables (x,y) will frequently be written as a partitioned column vector $\binom{x}{y}$. Gradients will always be regarded as column vectors. The notation $\nabla_c f$, for example, denotes the gradient of $f(c;x,y)$ with respect to c. All vectors of data are row vectors.

Vectors are denoted by lower case letters and matrices by upper case letters. A "bar" above a data symbol, e.g., $\overline{b}$, denotes data that are held fixed while other data are treated as parameters.

CHAPTER 9

SENSITIVITY TO OBJECTIVE FUNCTION DATA

Fluctuations in objective function data cannot result in a discrete jump in the optimal value. The sensitivity issue of relevance here is "how much" or "at what rate" can the optimal value change due to cost data changes. Although it is conceivable that this rate of change may be high, it is not typical that a small change in the "cost" data will result in a disproportionate change in optimal value.

Studying the sensitivity of v to cost data c consists of bounding and approximating $\bar{v}(c)$ over the range of values c of interest. An effective approach to approximating $\bar{v}(c)$ is to consider parametric cost data and address v in terms of the scalar parameter θ.

Procedures for solving linear problems parametric in the cost data are by now well documented. Noltemeier [13] presents basic properties and schema; a procedure in the context of parametric programming is described and computational results reported in the recent thesis by Nauss [12].

The intent here is to bring together the major facts already known about the behavior of v with respect to c and to show how these derive easily from a few fundamental properties. The optimization problem is given as:

$$\text{minimize } f(c;x,y) \ . \qquad (9.1)$$
$$(x,y) \epsilon F$$

A fundamental property of this problem is that fluctuations in the data c do not affect feasibility of a solution (x,y).

This observation suggests that the behavior of $\bar{v}(c)$ is founded on that of f with respect to c. First, the following upper bound is

131

immediate:

$$\overline{v}(c) \leq f(c;x,y) \qquad (9.2)$$

for any $(x,y) \in F$. Second, $\overline{v}(c)$ is monotone nondecreasing in c if f is monotone nondecreasing in c for each $(x,y) \in F$. Third, $\overline{v}(c)$ is continuous and concave (cf. Theorem 5.4) -- the only requirement for concavity is that f likewise be concave in c for $(x,y) \in F$ fixed.$\underline{*/}$ Thus, continuity and concavity of $\overline{v}$ hold even when f is discontinuous with respect to (x,y). Moreover, there is no requirement that F be a subset of a vector space: for instance, y may represent a combinatorial object such as a map of one finite set into another.

The bound (9.2) leads to another property of the nonlinear, possibly discrete problem (9.1), a property that is familiar in the context of LP problems. Suppose that the objective function is given by

$$f(c;x,y) \equiv \Sigma\ c_i h_i(x,y) \qquad (9.3)$$

where $h_i(x,y) \geq 0$ for all $(x,y) \in F$ so that $\overline{v}(c)$ is monotone nondecreasing. Define also

$$\underline{h_i} \equiv \min_F\ h_i(x,y), \quad \overline{h_i} \equiv \max_F\ h_i(x,y) \qquad (9.4)$$

and let (x^*,y^*) denote an optimal solution in problem (9.1).

Theorem 9.1: Let c' be any cost vector satisfying

$$c_i' \begin{cases} \leq c_i, & \text{if } h_i(x^*,y^*) = \overline{h_i} \\ \geq c_i, & \text{if } h_i(x^*,y^*) = \underline{h_i} \\ = c_i, & \text{otherwise .} \end{cases}$$

_____

$\underline{*/}$ Concavity of f in c generally precludes exponential data in the forms c^x, x^c, e^{cx} and log cx; however, many practical objective functions are concave (linear, in fact) in the data: polynomials on $X \times Y$, $\Sigma c_i e^{x_i}$, $\Sigma \ln(x_i)^{c_i}$, piecewise-linear concave functionals, etc.

Then (x^*, y^*) is optimal at c'.

Proof: Define

$$c_i'' = \begin{cases} c_i', & \text{if } h_i(x^*,y^*) = \underline{h}_i \\ c_i, & \text{otherwise} . \end{cases}$$

Without loss of generality assume that $\underline{h} = 0$ (i.e., translate: $h' = h-\underline{h}$). Then $\bar{v}(c'') \geq \bar{v}(c) = f(c;x^*,y^*) = f(c'';x^*,y^*) \Rightarrow (x^*,y^*)$ optimal at c''. Now $c_i' \leq c_i''$ if $h_i(x^*,y^*) = \bar{h}_i$ and $c_i' = c_i''$ otherwise. Translate: $h' \equiv h-\bar{h} \leq 0 \; [\Rightarrow \bar{v}(c') \geq \bar{v}(c'')]$. Then $f(c';x^*,y^*) = f(c'';x^*,y^*) = \bar{v}(c'') \leq \bar{v}(c') \Rightarrow (x^*,y^*)$ optimal at c'. $\quad\Vert$

In the special case where (9.1) is a linear MIP problem, the bounds (9.4) correspond to lower and upper bounds on the variables. Nauss [12] shows for the linear 0-1 MIP case that c' can be allowed even greater freedom when varying only one component at a time.

A fundamental property of (9.1) in the linear case is that

$$\min_{F} \; c\binom{x}{y} = \min_{Co(F)} \; c\binom{x}{y} \tag{9.5}$$

where, of course, the latter is an LP problem. It follows that $\bar{v}(c)$ is piecewise-linear and concave, just as in the LP case (Manne [11], Noltemeier [13]). If F is bounded, $\bar{v}(c)$ consists of a finite number of linear "pieces," $c \in E^n$.

Consider parametric costs $\bar{c}+\theta\bar{p}$. The concavity property permits a straightforward procedure for approximating $\hat{v}(\theta)$ [cf. (8.2)]. Assuming that f is differentiable with respect to c in direction $\bar{p}$ at $\bar{c}+\theta_0\bar{p}$, the bound (9.2) can be written as

$$\hat{v}(\theta) \leq L_0(\theta) \equiv \hat{v}(\theta_0) + (\theta-\theta_0)\bar{p}\cdot\nabla_c f(\bar{c};\bar{x},\bar{y}) \tag{9.6}$$

where $(\bar{x},\bar{y})$ is optimal at θ_0. The product $\bar{p}\cdot\nabla_c f$ represents the slope of

the linear upper bound. In the linear case (9.6) is simply $\hat{v}(\theta) \le (\overline{c} + \theta \overline{p})\,(\frac{\overline{x}}{\overline{y}})$.

Concavity also provides a means for lower-bounding: for any θ satisfying $\theta_1 \le \theta \le \theta_2$ where $\theta_1 < \theta_2$,

$$\hat{v}(\theta) \ge \underline{L}(\theta) \equiv \hat{v}(\theta_1) + (\theta - \theta_1)\left[\frac{\hat{v}(\theta_2) - \hat{v}(\theta_1)}{\theta_2 - \theta_1}\right] . \qquad (9.7)$$

By solving at the values θ_1 and θ_2 one obtains the bounds depicted in Figure 9.1. The shaded region represents uncertainty in the true graph of $\hat{v}(\theta)$. The concavity of $\hat{v}(\cdot)$ assures that $\hat{v}(\theta)$ does not behave wildly within the region of uncertainty: the rate of change (slope) of $\hat{v}(\theta)$, $\theta_1 \le \theta \le \theta_2$, is bounded by the slopes of $L_1(\theta)$ and $L_2(\theta)$.

An obvious procedure for reducing this uncertainty to an acceptable size is to resolve at an intermediate point such as θ_I, the point of intersection of $L_1(\theta)$ and $L_2(\theta)$. This introduces another upper bound and replaces the lower-bounding segment by two lower-bounding segments as shown in Figure 9.2.

Any procedure for approximating a concave function can be applied to $\hat{v}(\theta)$. A flexible procedure is offered in Appendix 9A that derives piecewise-linear concave functions $\tilde{V}(\theta)$ and $\tilde{v}(\theta)$ on [0,1] satisfying

$$\tilde{V}(\theta) \ge \hat{v}(\theta) \ge \tilde{v}(\theta)$$

and
$$\tilde{V}(\theta) - \tilde{v}(\theta) \le \epsilon$$

for any prescribed $\epsilon > 0$.
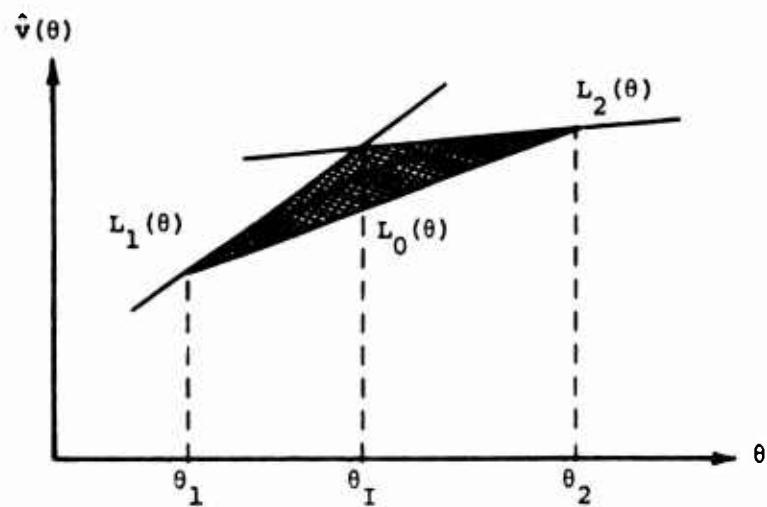
Figure 9.1.   Upper and Lower Bounds on $\hat{v}(\theta)$
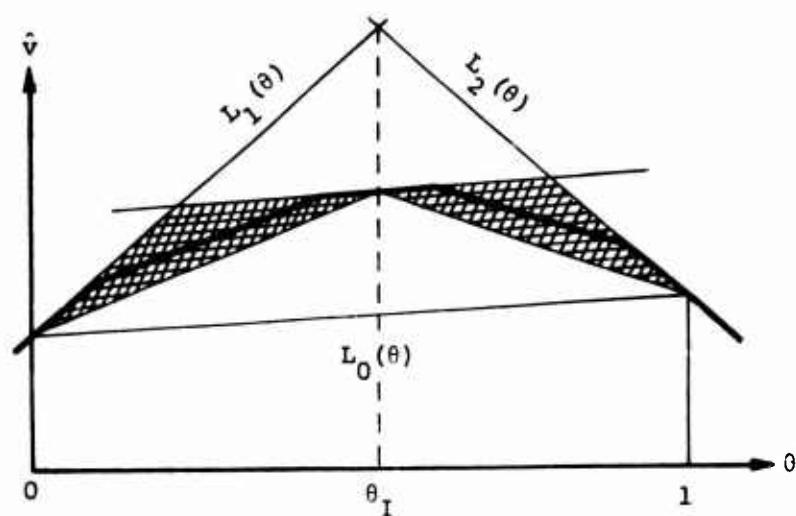
Figure 9.2.   Bounding Procedure for Deriving/Approximating $\hat{v}(\theta)$

A GENERAL PROCEDURE FOR APPROXIMATING $v(\theta)$

A basic procedure is described that applies in general to nonlinear MIP problems parametric in the cost data. This procedure specializes easily to the linear case.

Let $\varepsilon > 0$ denote a specified maximum allowable tolerance between $\hat{v}(\theta)$ and its approximation. The procedure will produce concave piecewise linear functions $\tilde{V}(\theta)$ and $\tilde{v}(\theta)$ satisfying

$$\tilde{V}(\theta) \geq \hat{v}(\theta) \geq \tilde{v}(\theta)$$

and

$$\tilde{V}(\theta) - \tilde{v}(\theta) \leq \varepsilon$$

(9A.1)

for all $\theta \in [0,1]$.

The maximum pointwise error in the estimate of $\hat{v}(\cdot)$ between θ_1 and θ_2 is given by the maximum difference between the upper bounds and the lower bound. This maximum difference, which will be denoted by h, occurs at the point θ_I where the upper bounds $L_1(\theta)$ and $L_2(\theta)$ intersect. Define $V_I \equiv L_1(\theta_I) = L_2(\theta_I)$. Thus,

$$h = V_I - \underline{L}(\theta_I)$$

(9A.2)

where θ_I is obtained by solving $L_1(\theta) = L_2(\theta)$ for θ.

If $h \leq \varepsilon$, then $\hat{v}(\theta)$ is sufficiently well approximated by

$$\tilde{V}(\theta) = \begin{cases} L_1(\theta), & \theta \in [\theta_1, \theta_I] \\ L_2(\theta), & \theta \in [\theta_I, \theta_2] \end{cases}$$

and

$$\tilde{v}(\theta) = \underline{L}(\theta), \quad \theta \in [\theta_1, \theta_2] \ .$$

(Initially, $\theta_1 = 0$ and $\theta_2 = 1$.) If $h > \varepsilon$, solve the MIP problem at θ_I. This yields $\hat{v}(\theta_I)$ and the linear support at θ_I:

$$\hat{L}(\theta) = \hat{v}(\theta_I) + (\theta - \theta_I)\hat{s} \tag{9A.3}$$

where $\hat{s} \equiv \overline{p} \cdot \nabla_c f(\overline{c} + \theta_I \overline{p}; x_I, y_I)$ and (x_I, y_I) is optimal at θ_I. The supports $L_1(\theta)$ and $L_2(\theta)$ can be expressed as

$$L_1(\theta) = V_I + (\theta - \theta_I)s_1$$
$$\tag{9A.4}$$
$$L_2(\theta) = V_I + (\theta - \theta_I)s_2$$

where s_1 and s_2 are the slopes given by

$$s_1 \equiv \overline{p} \cdot \nabla_c f(\overline{c} + \theta_1 \overline{p}; x_1, y_1)$$
$$s_2 \equiv \overline{p} \cdot \nabla_c f(\overline{c} + \theta_2 \overline{p}; x_2, y_2) \ .$$

These linear supports are depicted in Figure 9A.1. The supports $L_1(\theta)$ and $\hat{L}(\theta)$ intersect at θ_ℓ; $L_2(\theta)$ and $\hat{L}(\theta)$ intersect at θ_r. Two new lower-bounding line segments are generated. These are given by

$$\underline{L}^\ell(\theta) \equiv \hat{v}(\theta_I) + (\theta - \theta_I)\left[\frac{\hat{v}(\theta_I) - L_1(\theta_1)}{\theta_I - \theta_1}\right], \ \theta \ \epsilon \ [\theta_1, \theta_I]$$
$$\tag{9A.5}$$
$$\underline{L}^r(\theta) \equiv \hat{v}(\theta_I) + (\theta - \theta_I)\left[\frac{L_2(\theta_2) - v(\theta_I)}{\theta_2 - \theta_I}\right], \ \theta \ \epsilon \ [\theta_I, \theta_2] \ .$$

The maximum errors at θ_ℓ and θ_r are given, respectively, by

$$h_\ell = \hat{L}(\theta_\ell) - \underline{L}^\ell(\theta_\ell), \ h_r = \hat{L}(\theta_r) - \underline{L}^r(\theta_r) \ . \tag{9A.6}$$

If $s_1 = \hat{s}$, then $h_\ell = 0$. If $s_1 > \hat{s}$ (note that $s_1 \geq \hat{s} \geq s_2$), then it can be shown from (9A.3) through (9A.6) (or by observation from Figure 9A.1) that

137

$$\theta_\ell = \theta_I - \frac{V_I - \hat{v}(\theta_I)}{s_1 - \hat{s}}$$

(9A.7)

and

$$h_\ell = \left(\frac{\theta_\ell - \theta_1}{\theta_I - \theta_1}\right)[V_I - \hat{v}(\theta_I)] .$$

Similarly, if $s_2 = \hat{s}$, then $h_r = 0$, and if $s_2 < \hat{s}$, then

$$\theta_r = \theta_I + \frac{V_I - \hat{v}(\theta_I)}{\hat{s} - s_2}$$

(9A.8)

$$h_r = \left(\frac{\theta_2 - \theta_r}{\theta_2 - \theta_I}\right)[V_I - \hat{v}(\theta_I)] .$$
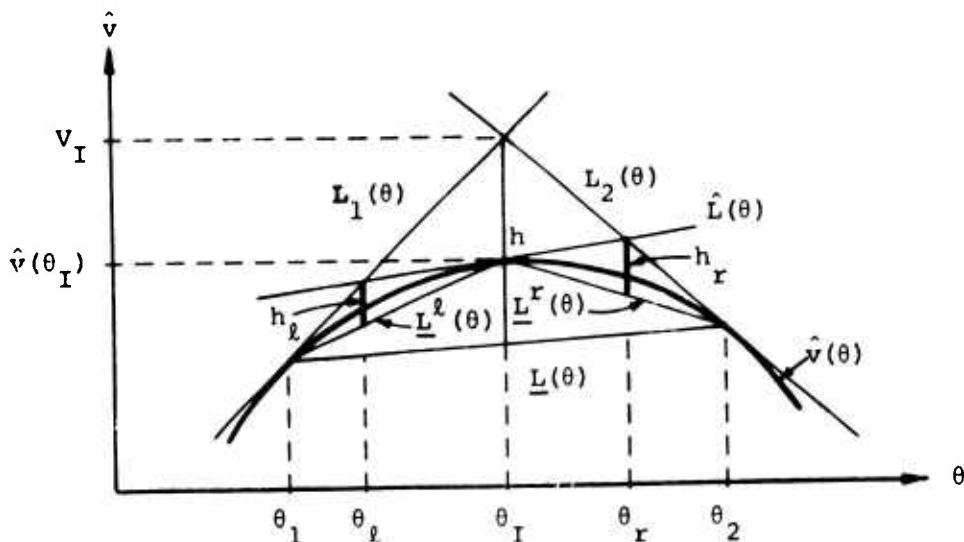
Figure 9A.1. The Approximation Procedure

The intersection point of two upper-bounding supports identifies what will be referred to as a candidate problem (CP). Thus, θ_ℓ and θ_r identify two candidate problems. Notice that h_ℓ, h_r and θ_ℓ, θ_r are expressed in terms of θ_I, V_I, s_1, s_2, θ_1 and θ_2. Consequently, a CP can be defined in terms of these six quantities and will be denoted by

$$CP:(\theta, V, s_1, s_2, \theta_1, \theta_2) .$$

If $h_\ell \leq \epsilon$, then $\hat{v}(\cdot)$ is sufficiently well approximated on $[\theta_1, \theta_I]$. If $h_\ell \geq \epsilon$, however, then the MIP problem to be solved at θ_ℓ is defined by

$$CP:(\theta_\ell, \hat{L}(\theta_\ell), s_1, \hat{s}, \theta_1, \theta_I) .$$

Similarly, $\hat{v}(\cdot)$ is sufficiently well approximated on $[\theta_I, \theta_2]$ if $h_r \leq \epsilon$; $h_r > \epsilon$ gives rise to the candidate problem

$$CP:(\theta_r, \hat{L}(\theta_r), \hat{s}, s_2, \theta_I, \theta_2) .$$

It is easily shown that

$$\hat{L}(\theta_\ell) = \frac{s_1 \hat{v}(\theta_I) - \hat{s}V_I}{s_1 - \hat{s}} \quad \text{and} \quad \hat{L}(\theta_r) = \frac{\hat{s}V_I - \hat{s}_2 \hat{v}(\theta_I)}{\hat{s} - s_2} . \tag{9A.9}$$

The procedure requires maintaining a list of candidate problems. Each time a CP is solved, this list is reduced by one and increased by zero, one, or two additional entries. The solution of each CP yields a point $(\theta, \hat{v}(\theta))$; each time the error h_I at an intersection point θ_I (i.e., θ_ℓ or θ_r) is less than or equal to ϵ, the information (θ_I, V_I, h_I) gives acceptable estimates of $\hat{v}(\cdot)$ at θ_I and the maximum error in these estimates -- specifically,

$$\tilde{v}(\theta_I) \equiv V_I \geq \hat{v}(\theta_I) \geq V_I - h_I = \tilde{v}(\theta_I) .$$

139

The procedure is described as follows:

(1) Solve for $\hat{v}(0)$ and $\hat{v}(1)$ and output the points $(0,\hat{v}(0))$, $(1,\hat{v}(1))$. Compute h given by (9A.2). If $h \leq \epsilon$, output (θ_I, V_I, h) and STOP.

(2) Solve at θ_I and output $(\theta_I, \hat{v}(\theta_I))$. Compute h_ℓ and h_r (at θ_ℓ and θ_r, respectively) from (9A.7) and (9A.8). If $h_\ell \leq \epsilon$, output $(\theta_\ell, \hat{L}(\theta_\ell), h_\ell)$; i.e., $\tilde{V}(\theta_\ell) = \hat{L}(\theta_\ell)$. If $h_\ell > \epsilon$, enter CP:$(\theta_\ell, \hat{L}(\theta_\ell), s_1, \hat{s}, \theta_1, \theta_I)$ into the CP list. If $h_r \leq \epsilon$, output $(\theta_r, \hat{L}(\theta_r), h_r)$; if $h_r > \epsilon$, enter CP:$(\theta_r, \hat{L}(\theta_r), \hat{s}, s_2, \theta_I, \theta_2)$ into the CP list.

(3) If the CP list is empty, STOP. Otherwise, select a CP:$(\theta_I, V_I, s_1, s_2, \theta_1, \theta_2)$ from the list and go to (2).

The procedure provides two sets of points which we will refer to as S_1 and S_2. The set S_1 contains points of the form $(\theta, \hat{v}(\theta))$ obtained by solving the candidate problems. At these values of θ, $\tilde{V}(\theta) = \hat{v}(\theta) = \tilde{v}(\theta)$. The set S_2 contains points of the form $(\theta, \tilde{V}(\theta))$ obtained by observing an acceptable error at points of intersection of linear supports. The lower-bounding function $\tilde{v}(\theta)$ is obtained by linear interpolation between adjacent points of S_1 (ordered by increasing values of θ); similarly, $\tilde{V}(\theta)$ is obtained by linear interpolation between adjacent points of $S_1 \cup S_2$.

These functions are illustrated in Figure 9A.2 with $S_1 = \{(0,v_0), (\theta_2,v_2), (1,v_1)\}$ and $S_2 = \{(\theta_3,v_3), (\theta_4,v_4)\}$. Associated with each point in S_2 is the computed maximum error; these are indicated as h_3 and h_4 ($h_3 \leq \epsilon$, $h_4 \leq \epsilon$).

140
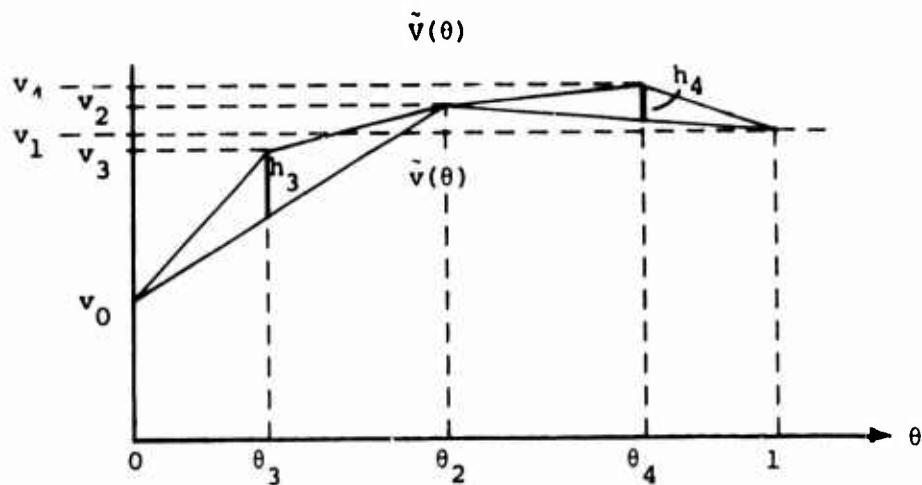
Figure 9A.2.   Illustration of the Bounding Functions $\tilde{V}$ and $\tilde{v}$

The procedure can be enhanced by using feasible solutions generated
in the course of solving a CP to define additional CPs. For instance,
solving at $\theta = 0$ in step (1) produces feasible solutions that are not
optimal at $\theta = 0$ but are possibly optimal at other values of θ. Taking
the pointwise minimum of the upper bounds associated with these solu-
tions defines additional CPs. Evidently, it is possible in the linear
case that this feature could derive $\hat{v}(\theta)$ after solving N+2 problems
where N is the number of break points (at least this many are required
to derive $\hat{v}(\theta)$).

The proof of finite convergence requires establishing that the
maximum errors generated by a chain of successive CPs tend toward zero.
A proof is given later. The rate of convergence depends, of
course, on the tolerance ε and the degree of nonlinearity of $\hat{v}(\theta)$. It is
worth mentioning that in practice ε should be chosen larger than any
tolerance on optimal (i.e., near-optimal) solution values; (the linear
supports, θ_I, V_I, and maximum errors h can be no more accurate than the
solutions from which they are determined).

141

For <u>linear</u> MIP problems $\hat{v}(\theta)$ is known to be piecewise linear (and concave). The approximation procedure can be enhanced by exploiting piecewise linearity. Each of the finitely-many break points of $\hat{v}(\theta)$ is an intersection point of two upper bounding linear supports. Since the approximation procedure selects such intersection points to define successive candidate problems, the procedure with $\varepsilon = 0$ would (in theory) eventually discover all break points and thereby produce $\hat{v}(\theta)$. Thus, finite convergence is evident for all $\varepsilon \geq 0$.

Two improvements, in particular, can be made in the approximation procedure when applied to linear MIP problems parametric in the cost data. Given $CP:(\theta_I, V_I, s_1, s_2, \theta_1, \theta_2)$, if $\hat{v}(\theta_I) = \underline{L}(\theta_I)$, then $\hat{v}(\theta) = \underline{L}(\theta)$, for all $\theta \in [\theta_1, \theta_2]$. It can be shown that $\underline{L}(\theta_I)$ can be expressed in terms of the six available quantities as follows:

$$\underline{L}(\theta_I) = V_I - (s_1 - s_2) \frac{(\theta_I - \theta_1)(\theta_2 - \theta_I)}{\theta_2 - \theta_1} \ . \qquad (9A.10)$$

Secondly, if $\hat{v}(\theta_I) = V_I$, then θ_I is a break point and $\hat{v}(\theta) = \min\{L_1(\theta), L_2(\theta)\}$ on $[\theta_1, \theta_2]$; this CP is then "fathomed" in the sense that no descendant CPs need be generated.

In practice, $\hat{v}(\theta_I)$ would be tested for approximate equality to V_I or $\underline{L}(\theta_I)$ --- $\hat{v}(\theta_I) \approx V_I$ or $\hat{v}(\theta_I) \approx \underline{L}(\theta_I)$ --- to provide for a computational tolerance on optimality and the propagation thereof in the computation of V_I and $\underline{L}(\theta_I)$. Notice that if $\hat{v}(\theta_I)$ is within ε of V_I, then the CP is fathomed since the procedure will find that $h_\ell < \varepsilon$ and $h_r < \varepsilon$.

These tests can be implemented in step (2) of the procedure before h_ℓ and h_r are computed. If either $\hat{v}(\theta_I) \approx V_I$ or $\hat{v}(\theta_I) \approx \underline{L}(\theta_I)$, then output the point $(\theta_I, \hat{v}(\theta_I))$ and go to step (3) -- the current CP is

fathomed.

Theorem 9A.1: Suppose that the MIP problem is linear; let N (N $\geq$ 1)
denote the number of break points in [0,1]. Then the procedure con-
verges after solving no more than 2N+1 problems (including θ = 0 and
θ = 1), for any $\varepsilon \geq 0$.

Proof: There are N+1 linear segments on the graph of $\hat{v}(\theta)$, $\theta \in [0,1]$.
Suppose θ_I is not a break point; let θ_1, θ_2 denote the adjacent break
points which identify the linear segment on which $(\theta_I, \hat{v}(\theta))$ lies. This
linear segment coincides with the linear support $\hat{L}(\theta)$ generated by the
solution at θ_I. Consequently, any other (upper-bounding) support cannot
intersect $\hat{L}(\theta)$ at a point $\theta \in (\theta_1, \theta_2)$. Therefore, no more than one
problem will be solved for each of the N+1 segments. Solving at each
of the N break points thus yields a maximum total of 2N+1 problems. ||

   We now give a proof of finite convergence for the nonlinear case.[*]
The solution of the first CP (candidate problem) gives rise to two
more "second generation" CPs. An nth generation CP results from solv-
ing n-1 predecessor CPs and can itself father zero, one, or two $(n+1)^{th}$
generation CPs. Let $\{(CP)^n\}$ denote a sequence of CPs where $(CP)^n$ is of
the nth generation, and let h^n be the computed maximum error correspond-
ing to $(CP)^n$. The problem $(CP)^n$ is identified by

$$(CP)^n : (\theta^n, v^n, s_1^n, s_2^n, \theta_1^n, \theta_2^n) \ .$$

Also, let s_0^n denote the slope of the lower bound for $(CP)^n$.

_____

[*] Noltemeier [13] gives an approximation approach similar to the one
described here and a tedious proof of convergence (no bound is given
on the number of problems that must be solved). The procedure des-
cribed here takes advantage of the slope generated at θ_I to impose an
improved upper bound (this bound also determines the descendant CPs)
and the proof of convergence is relatively easy.

143

<u>Lemma 1</u>: The sequence $\{h^n\}$ converges monotonically.

<u>Proof</u>: For all n, $h \geq h^n \geq 0$ where h is computed in step (1) -- see Figure 9A.1. Also, $h^n \geq h^{n+1}$ (by observing similar triangles in Figure 9A.1). Boundedness and monotonicity imply convergence. $\quad || $

Let $\bar{h} \equiv \lim_{n \to \infty} h^n$. If for some n, $h^n = 0$, then $\{(CP)^n\}$ is a finite sequence and $\bar{h} = 0$. Suppose on the other hand that $h^n > 0$ for all n. Define $p^n \equiv V^n - v(\theta^n)$ and $q^n \equiv h^n - p^n$. These quantities are illustrated in Figure 9A.3.
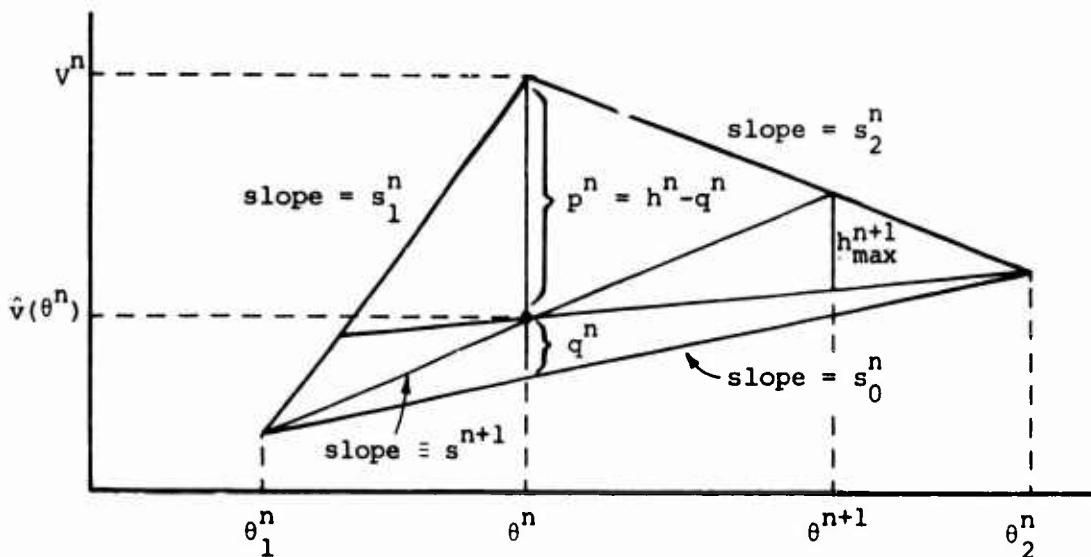
Figure 9A.3. Illustration of Descriptive Quantities for $(CP)^n$

<u>Lemma 2</u>: $h^n \to 0$.

<u>Proof</u>: Since $\bar{h} \leq h^{n+1} < p^n \equiv h^n - q^n < h^n$, we have

$$\bar{h} = \lim h^{n+1} = \lim h^n - \lim q^n = \bar{h} - \lim q^n \leq \bar{h}$$

so that $q^n \to 0$. This in turn implies that $|s^{n+1} - s_0^n| \to 0$ (these slopes are bounded by the finite slopes s_1 and s_2). Also,

144

$$0 \le h^{n+1} \le h_{max}^{n+1} < (\theta^{n+1} - \theta^n) \cdot |s^{n+1} - s_0^n|$$

$$\le |s^{n+1} - s_0^n| \to 0 . \qquad\qquad ||$$

<u>Theorem 9A.2</u>: The approximation procedure ($\varepsilon > 0$) converges in a finite number of steps.

<u>Proof</u>: For any chain $\{(CP)^n\}$, $h^n \to 0$. Hence, $\exists\ N \ni h^n \le \varepsilon\ \forall\ n \ge N$. The result follows since each chain is terminated after a finite number of generations, thereby permitting only a finite number of chains.

CHAPTER 10

RIGHT-SIDE ANALYSIS: BASIC CONCEPTS AND PROPERTIES

Methods for determining the behavior of the optimal value v with

respect to variations in right-side (RS) data derive from a handful of

basic concepts and properties. Fundamental are the restriction and

relaxation concepts which provide upper and lower bounds (respective-

ly) on v. The judicious combination of these upper and lower bounds

provides a means for analyzing the behavior of v as a function of RS

data.

Much of the development of this chapter makes use of recent works

by Geoffrion [4,5,6] and some familiarity with these will be helpful.

The material is presented in the context of nonparametric RS analysis,

but is easily specialized to the parametric case $(b+\theta r)$.

10.1 The Analysis Problem

We address here the function $\overline{v}(b)$ defined as the optimal value of

the MIP problem

$$(P) \begin{cases} \text{Minimize } f(\overline{c};x,y) \\ (x,y) \epsilon U \\ \text{subject to: } G(\overline{a};x,y) \leq b \end{cases}$$

where it is assumed that the data $\overline{c}$ and $\overline{a}$ are fixed, $b \epsilon E^m$, and

$U \equiv C \cap (X \times Y)$ where X is compact in E^n $(x \epsilon E^n)$, Y is a finite set in

E^q $(y \epsilon E^q)$, and C is a closed set in E^{n+q}. The (compact) feasible

region $\{(x,y) \epsilon U | G(\overline{a};x,y) \leq b\}$ will be denoted by $F(b)$. The set of

right-sides for which (P) is feasible is denoted by B; $\overline{v}(b) \equiv \infty$ if

$b \notin B$.

It is assumed also that f is continuous on $\overline{c} \times X @ Y$, G is continu-
ous on $\overline{a} \times X @ Y$, and $U \neq \emptyset$. To simplify notation slightly, $f(\overline{c};x,y)$
and $G(\overline{a};x,y)$ henceforth will be written as $f(x,y)$ and $G(x,y)$.

Formulation (P) suggests a partitioning of the problem constraints
into two sets: those with right-sides that are not permitted to be
perturbed (represented by the set U) and those with right-side b whose
influence on v is to be investigated. Each of these sets may include
equality as well as inequality constraints.

The analysis problem is to <u>determine right-sides $\hat{b}$</u> that are accept-
ably "close" to the nominal b and <u>which grant (net) improvements in</u>
<u>optimal value</u>; (an optimal solution at such a point $\hat{b}$ is referred to as
an <u>alternative</u> solution to (P)). With parametric RS, $\overline{b}+\theta\overline{r}$, an additional
objective may be to derive or approximate $v(\theta)$.

10.2 Restriction and Upper Bounding

The term restriction is used in the following sense: a minimiza-
tion problem (Q) is said to be a <u>restriction</u> of (P) if for all $b \in B$
the feasible region F(b) of (P) contains that of (Q) and the objective
function of (Q) is nowhere less than $f(\cdot,\cdot)$ on F(b). The idea of upper
bounding by restriction is a familiar one and is briefly reviewed here
in the context of right-side analysis.

Restricting any subset of the variables (x,y) to fixed feasible
values leads to an upper bound on $\overline{v}(b)$. In particular, when the dis-
crete variables are held fixed at a $y \in Y$, one has

$$\overline{v}(b) \leq \overline{v}_y(b), \quad (y \in Y, \, b \in E^m) . \tag{10.1}$$

The problem (P_y) obtained by fixing y in (P) is a continuous-variable
problem. When C is a convex set and f and G are convex in x, $\overline{v}_y(\cdot)$ is

147

also convex.

A natural choice of y in (10.1) is a y^* that is optimal at $\bar{b}$ in (P);
thus, equality holds in (10.1) at $b = \bar{b}$ and it is reasonable to expect
that equality holds on a larger subset of B. This conjecture is intui-
tive from the fact that

$$\bar{v}(b) = \min_{y \in Y}\{\bar{v}_y(b)\}, \quad b \in B \tag{10.2}$$

and that B can be partitioned into a finite number of subsets on each of
which $\bar{v}(\cdot) = \bar{v}_y(\cdot)$ for some $y \in Y$. The following properties suggest
that $\bar{v}_{y^*}(\cdot)$ is a reliable indicator of the <u>local</u> behavior of $\bar{v}(\cdot)$.

(1) For any $\varepsilon > 0$ there is a $\Delta b > 0$ such that

$$\bar{v}_{y^*}(b) - \varepsilon \le \bar{v}(b) \le \bar{v}_{y^*}(b), \quad \text{for } \bar{b} \le b \le \bar{b}+\Delta b . \tag{10.3}$$

(2) If y^* is strongly feasible (this precludes equality constraints
in $G(x,y) \le b$; cf. §4.3), then for any $\varepsilon > 0$ there exists a
$\Delta b' > 0$ such that

$$\bar{v}_{y^*}(b) - \varepsilon \le \bar{v}(b) \le \bar{v}_{y^*}(b), \quad \text{for } \bar{b}-\Delta b' \le b \le \bar{b}+\Delta b' . \tag{10.4}$$

(3) If y^* is uniquely optimal at $\bar{b}$, then:

(a) there exists a $\Delta b > 0$ such that

$$\bar{v}(b) = \bar{v}_{y^*}(b) \quad \text{for } \bar{b} \le b \le \bar{b}+\Delta b ; \tag{10.5}$$

(b) if in addition y^* is strongly feasible at $\bar{b}$ (all con-
straints $G(x,y) \le b$ must be inequalities), then there
exists a $\Delta b' > 0$ such that

$$\bar{v}(b) = \bar{v}_{y^*}(b) \quad \text{for } \bar{b}-\Delta b' \le b \le \bar{b}+\Delta b' . \tag{10.6}$$

Property (1) derives easily from lower-semicontinuity of $\bar{v}$ and
continuity of $\bar{v}_{y^*}$. The choice of b assures that y^* is feasible in (P).
Strong feasibility assures that y^* is feasible in a neighborhood of $\bar{b}$,

148

thus property (2) results as a simple extension of (1).

A detailed proof of property (3) is given in Appendix 10A along with a brief discussion on the use of dual variable interpretations in approximating local behavior. Under the hypotheses of property (3), the MIP problem (P) locally reduces to a continuous-variable problem; the usual interpretation of optimal duals corresponding to $\bar{b}$ in problem (P_{y^*}) is equally valid when applied to (P). Unfortunately, the size of the neighborhood of $\bar{b}$ on which $\bar{v}$ and $\bar{v}_{y^*}$ coincide is not known. These remarks hold, of course, for pure integer problems as well as for MIP problems.

Upper bounding by monotonicity and feasibility are elementary:

$$\bar{v}(b) \leq \bar{v}(\bar{b}) \qquad \text{for all } b \geq \bar{b}$$

$$\bar{v}(b) \leq f(x,y) \quad \text{for all } b \in B \text{ such that}$$
$$(x,y) \text{ is feasible in (P) .}$$

10.3 Relaxations and Lower Bounding

Lower bounds on $\bar{v}(b)$ derive from relaxations of (P). Linear lower bounding functions are easily derived from a single solution of a (continuous-variable) relaxation and can point out the possibility or impossibility of substantial jumps in the optimal value of $\bar{v}(b)$.

A minimizing problem (Q) is said to be a relaxation of (P) if for any $b \in B$ the feasible region of (Q) contains $F(b)$ and the objective function of (Q) nowhere exceeds $f(\cdot,\cdot)$ on $F(b)$. This definition accommodates a generalized Lagrangean program*/ within its scope.

A common approach to relaxation of (P) is convexification in the domain space. Let W be any convex set in E^{n+q} such that $U \subseteq W$. (Since

---
*/See, e.g., Roode [15].

149

U is compact, it can be assumed without loss of generality that W is compact.) The continuous-variable problem

$$(P_W): \quad \text{Minimize } f(x,y) \,|\, G(x,y) \leq b \\ (x,y) \epsilon W$$

is a relaxation of (P). Of particular interest will be the sets $W \equiv Co(U)$ and $W \equiv U_0$ where U_0 denotes U without the integrality restrictions on y. The optimal value of (P_W) is denoted by $\overline{v}_W(b)$. It will be assumed in this section that f and G are convex on W. Hence, $\overline{v}_W(\cdot)$ is also convex.

<u>Theorem 10.1</u>: Given that (P_W) has a feasible solution, if μ ($\mu \leq 0$) is a vector of optimal dual variables[*] for (P_W), then

$$\overline{v}(b+\Delta b) \geq \overline{v}_W(b) + \mu \cdot \Delta b \qquad (10.3)$$

for all $\Delta b \epsilon E^m$.

<u>Proof</u>: The fact that μ is a subgradient to $\overline{v}_W(\cdot)$ at b follows from [5, Th. 1]; thus,

$$\overline{v}_W(b+\Delta b) \geq \overline{v}_W(b) + \mu \cdot \Delta b \qquad (10.4)$$

and (10.3) follows since (P_W) is a relaxation of (P). ||

This result clearly holds also when the objective function of (P_W) is any convex function on W which nowhere exceeds $f(\cdot,\cdot)$ on F(b).

A comment on the existence of μ is in order. Let $B_W \equiv \{b \epsilon E^m |$ (P_W) is feasible}. A vector of optimal dual variables exists (is finite)

---

[*] That is, μ satisfies the usual optimality conditions--see Geoffrion [5], p. 5--with the convention that $\mu \leq 0$. This convention is adopted so that μ can be interpreted as a subgradient of $\overline{v}_W(\cdot)$ without the need for a sign change.

for (P_W), $b \in B_W$, if (P_W) is "stable":[*] there exists $M > 0$ such that $\overline{v}_W(b) - \overline{v}_W(b+\Delta b) \leq M\|\Delta b\|$ for all $\Delta b \in E^m$ (where the norm $\|\cdot\|$ is arbitrary; note that $\overline{v}_W(b)$ is finite for $b \in B_W$ since W is compact). Convexity and finiteness of $\overline{v}_W(\cdot)$ on B_W assures that an optimal μ exists when $b \in \text{int}(B_W)$ -- (cf. [5, Th. 6]). If b is a boundary point of B_W, (P_W) is clearly stable if the objective function f satisfies a Lipschitz condition: there exists an $M > 0$ such that

$$\left| f(x,y) - f(x',y') \right| \leq M\|(x,y) - (x',y')\|$$

for all $(x,y), (x,',y') \in W$. If (P_W) is linear, μ exists and is available from the final tableau of the simplex procedure.

Combining the lowe· bound (10.3) with upper bounding by feasibility results in the following ε-optimality property.

Theorem 10.2: Let $(\overline{x},\overline{y})$ and μ be optimal (primal and dual variables, respectively) in problem (P_W). Let (x,y) and Δb be such that (x,y) is feasible in (P) with right-side $b+\Delta b$. Then (x,y) is ε-optimal at $b+\Delta b$ in (P) where $\varepsilon \equiv \Delta f - \mu \cdot \Delta b$ ($\Delta f \equiv f(x,y) - f(\overline{x},\overline{y}) = f(x,y) - \overline{v}_W(b)$).

Proof:
$$0 \leq f(x,y) - \overline{v}(b+\Delta b) \quad \text{by feasibility}$$
$$\leq f(x,y) - \overline{v}_W(b) - \mu \cdot \Delta b \quad \text{by (10.3)}$$
$$= f(x,y) - f(\overline{x},\overline{y}) - \mu \cdot \Delta b$$
$$\equiv \varepsilon . \qquad\qquad ||$$

An obvious application of Theorem 10.2 is to the "rounding" of a $\overline{y}$ which, together with $\overline{x}$, is optimal in (P_W). This application is studied in Chapter 11; it will be shown that for certain specializations of (P), $\varepsilon = 0$ in Theorem 10.2 when y is selected as a rounding of $\overline{y}$. It is

_____

[*] Cf. Gale [2], Geoffrion [5].

151

worth noting that a nondiscrete $\overline{y}$ may be acceptable in a quasi-integer sense (cf. Ch. 7).

It will be instructive to study the properties of problem (P_W) where $W \equiv Co(U)$. This problem is denoted by (P_*) and its optimal value by $\overline{v}_*(b)$:

$$(P_*) \quad \underset{(x,y) \in Co(U)}{\text{Minimize}} \quad f(x,y) \mid G(x,y) \leq b \ .$$

The following result is a generalization of a property established by Geoffrion [6, Th. 4].

<u>Theorem 10.3</u>: The (convex) function $\overline{v}_*(\cdot)$ is the lower convex envelope of $\overline{v}(\cdot)$.

<u>Proof</u>:
$$E \equiv \text{Epigraph}[\overline{v}(\cdot)] = \{(\lambda,b) \mid \lambda \geq \overline{v}(b)\}$$
$$= \{(\lambda,b) \mid \lambda \geq f(x,y), \ G(x,y) \leq b \text{ for some } (x,y) \in U\}.$$

E^* is defined in the same way except that U is replaced by $Co(U)$. The theorem asserts that E^* is the convex hull of E. The epigraph E^* is convex since v_* is convex. Also, $\overline{v}_*(b) \leq \overline{v}(b) \ \forall b$ so that $E^* \supseteq Co(E)$. Next, suppose that $(\overline{\lambda},\overline{b}) \in E^*$. Then $\overline{\lambda} \geq f(\overline{x},\overline{y})$ and $G(\overline{x},\overline{y}) \leq \overline{b}$, for $(\overline{x},\overline{y}) \in Co(U)$. There exists $(x^j,y^j) \in U$ such that $(\overline{x},\overline{y}) = \Sigma\beta^j(x^j,y^j)$, $(\beta^j \geq 0, \ \Sigma\beta^j = 1)$. Define $\lambda^j \equiv f(x^j,y^j)$, $b^j \equiv G(x^j,y^j)$. Clearly, $(\lambda^j,b^j) \in E \ \forall j$. Hence, $\Sigma\beta^j(\lambda^j,b^j) \in Co(E)$ and

$$\Sigma\beta^j(\lambda^j,b^j) = (\Sigma\beta^j f(x^j,y^j), \ \Sigma\beta^j G(x^j,y^j))$$
$$\leq (f(\overline{x},\overline{y}), \ G(\overline{x},\overline{y})) \quad \text{by convexity of } f \text{ and } G$$
$$\leq (\overline{\lambda},\overline{b})$$

which implies that $(\overline{\lambda},\overline{b}) \in Co(E)$ and $E^* \subseteq Co(E)$. $\mid\mid$

When $W \equiv Co(U)$ the lower bound relationship (10.3) holds with equality for some perturbation Δb to the right-side b:

152

<u>Corollary 10.3.1</u>: Let μ be an optimal multiplier vector (subgradient) in (P_*). There exists a Δb such that

$$\overline{v}(b+\Delta b) = \overline{v}_*(b) + \mu \cdot \Delta b = \overline{v}_*(b+\Delta b) . \qquad (10.5)$$

<u>Proof</u>: Suppose not. Then $\overline{v}(b+\Delta b) > \overline{v}_*(b) + \mu \cdot \Delta b$ for all Δb. Since $X \times Y$ is compact, all values for $\overline{v}(\cdot)$ and $\overline{v}_*(\cdot)$ are generated on a compact domain B'. Hence, there exists a $\gamma > 0$ such that $\overline{v}(b+\Delta b) \geq \overline{v}_*(b) + \mu \cdot \Delta b + \gamma \equiv \Gamma(\Delta b)$. The pointwise maximum of $\Gamma(\Delta b)$ and $\overline{v}_*(b+\Delta b)$ is a convex lower bound on $\overline{v}(b+\Delta b)$ which exceeds $\overline{v}_*$ by γ at $\Delta b = 0$, contradicting the fact that $\overline{v}_*$ is the envelope function. $\qquad ||$

The corollary implies that $\overline{v}_*(b) + \mu \cdot \Delta b$ is a linear support to $\overline{v}(\cdot)$ when $W = Co(U)$, which in turn implies that (x,y) and Δb in Theorem 10.2 can be selected so that $\epsilon = 0.$[*] Consequently, among all functions linear in Δb, ϵ as defined in Theorem 10.2 gives the smallest measure of suboptimality (when $W = Co(U)$).

Geometrically, the corollary implies that $\overline{v}_*(\cdot)$ is linear on regions of B where $\overline{v}(b) > \overline{v}_*(b)$ and that $\overline{v}(\hat{b}) = \overline{v}_*(\hat{b})$ at all extreme points $\hat{b}$ of (the epigraph of) $\overline{v}_*$. See Figure 10.1. This property of the envelope function $\overline{v}_*$ is intuitive and holds for an arbitrary l.s.c. function $\overline{v}(\cdot)$; a detailed proof is given in Appendix 10B.

---

[*] By Corollary 10.3.1 there exist (x,y) and Δb such that
$$0 = \overline{v}(b+\Delta b) - [\overline{v}_*(b)+\mu \cdot \Delta b] = [f(x,y)-\overline{v}_*(b)] - \mu \cdot \Delta b = \epsilon.$$
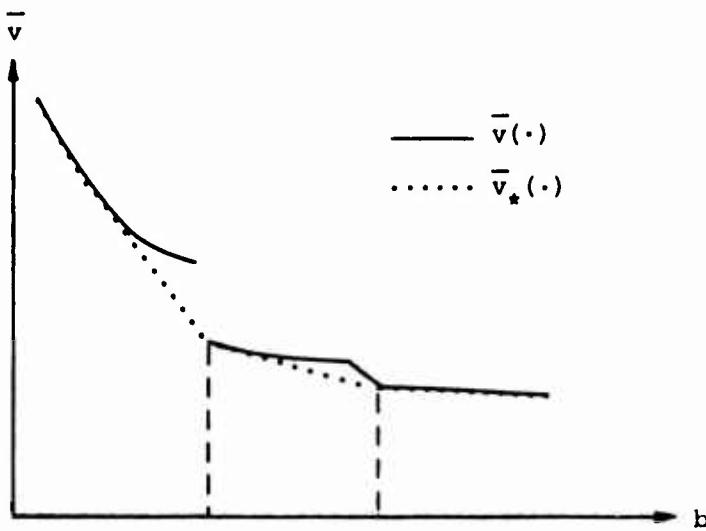
Figure 10.1. Linearity of $\overline{v}_*$ where $\overline{v}_* < \overline{v}$, and
Equality of $\overline{v}_*$ and $\overline{v}$ at Extreme Envelope Points

In general it may be difficult to characterize $\text{Co}(U)$ as an explicit
set of constraints, but this is possible at least in principle. In
certain special cases of (P), $\text{Co}(U)$ coincides with the convex set ob-
tained by simply dropping the integrality constraints on y. To wit,
let Y_0 be such that $Y = \{y \in Y_0 | y \text{ integer}\}$ and suppose that $U = X \times Y$;
if $Y_0 = \text{Co}(Y)$ then $U_0 \equiv X \times Y_0 = \text{Co}(U).\underline{*/}$ This clearly holds, for
example, when $Y \equiv \{y | 0 \le y_i \le u_i, \; y_i \text{ integer}\}$, and also holds when Y
includes GUB constraints on y as well as bounds and integrality.$\underline{\dagger/}$
In these special cases $\overline{v}_*(\cdot) = \overline{v}_0(\cdot)$, where $\overline{v}_0(\cdot)$ corresponds to $W \equiv U_0$
in (P_W).

$\underline{*/}\text{Co}(U) = \text{Co}(X \times Y) = \text{Co}(X) \times \text{Co}(Y) = X \times \text{Co}(Y).$
$\underline{\dagger/}$ Let $Y \equiv \{y \in Y_0 | y \text{ integer}\}$, $Y_0 \equiv \{y | y_{ij} \ge 0, \Sigma_j y_{ij} = 1, (i = 1,\ldots,I)\}$.
Since Y_0 is the cross-product of sets $Y_0^i \equiv \{y | 0 \le y_j \le 1, \Sigma y_j = 1\}$ it
suffices to show that $Y_0^i = \text{Co}(Y^i)$. $\text{Co}(Y^i) \subseteq Y_0^i$ since Y_0^i is convex. Let
$y \in Y_0^i$ and write $y = \Sigma y_j e_j$ where $e_j =$ unit vector with jth component
unity. Each $e_j \in Y^j$ so $y \in \text{Co}(Y^i)$ and $Y_0^i \subseteq \text{Co}(Y^i)$.

154

10.4 A General Relaxation

A relaxation is studied here which permits the "costs" of additional resources Δb to be represented explicitly in the model. This relaxation of (P) produces, besides a linear lower bounding function, an alternative solution to (P) that nets the greatest improvement in optimal value.

Let $p(\cdot)$ be any function $p:E^m \to E$ satisfying $p(0) = 0$ and $p(z) \leq 0$ for $z \leq 0$. Such functions will be termed "admissible." The following problem is a relaxation of (P):

$$(G) \quad \underset{(x,y)\epsilon U}{\text{Minimize}} \; f(x,y) + p(G(x,y)-b) \equiv V(p) \; .$$

The constraint set U can be replaced by any set $W \supseteq U$; in either case one has

$$\bar{v}(b) \geq V(p) \tag{10.6}$$

for all admissible functions p. Several additional relationships between (P) and (G) suggest the use of formulations similar to (G) for the purpose of discovering perturbations Δb to b that yield net improvements in optimal value.

<u>Theorem 10.4</u>: (a) Given $p(\cdot)$, if $(\hat{x},\hat{y})$ satisfies the conditions

(i) $(\hat{x},\hat{y})$ is optimal in (G)

(ii) $G(\hat{x},\hat{y}) \leq b$

then $(\hat{x},\hat{y})$ is ϵ-optimal in (P) where $\epsilon \equiv -p(G(\hat{x},\hat{y})-b)$.

(b) If $(\hat{x},\hat{y})$ is optimal in (G), then $(\hat{x},\hat{y})$ is optimal in (P) with b replaced by $\hat{b} \equiv G(\hat{x},\hat{y})$, and

$$\bar{v}(\hat{b}) = V(p) - p(\hat{b}-b) \; . \tag{10.7}$$

(c) $\qquad\qquad \bar{v}(b+\Delta b) \geq V(p) - p(\Delta b)$ for all $\Delta b \; \epsilon \; E^m$. $\qquad$ (10.8)

155

<u>Proof</u>: Let $\hat{b} \equiv G(\hat{x},\hat{y})$. By definition, $V(p) =$

$$f(\hat{x},\hat{y})-\epsilon \leq \overline{v}(b) \qquad \text{[by (10.6)]}$$
$$\leq f(\hat{x},\hat{y}) \quad \text{[by feasibility of } (\hat{x},\hat{y})$$
$$\text{in problem (P)] .}$$

This shows part (a). To establish (b) and (c), let $\Delta b = G(x,y)-b$ and project (G) onto (the variables) Δb:

$$V(p) = \min_{\Delta b} [p(\Delta b) + \min_{(x,y)\epsilon U} f(x,y)\,|\,G(x,y) = b+\Delta b]$$

$$= \min_{\Delta b} [p(\Delta b) + \overline{v}(b+\Delta b)]$$

since the inner problem is (P) with b replaced by b+Δb. Clearly, $V(p) \leq p(\Delta b) + \overline{v}(b+\Delta b)$ for all Δb; if $\hat{\Delta b}$ is optimal in the outer problem, then $V(p) = p(\hat{b}-b) + \overline{v}(b)$ where $\hat{b} \equiv b+\hat{\Delta b} = G(\hat{x},\hat{y})$. ||

The theorem holds for completely general objective and constraint functions and all admissible functions $p(\cdot)$. Notice also that Theorem 10.4(c) holds when U in (G) is replaced by any convex set $W \supseteq U$.

Solving problem (G) produces a solution that is optimal in (P) at some perturbed right-side b+$\hat{\Delta b}$ and that provides the **maximum** net improvement in optimal value.[*] Conditions (i) and (ii) together with the requirement that $p(\hat{\Delta b}) = 0$ are sufficient conditions for a solution $(\hat{x},\hat{y})$ to be optimal in (P). If p satisfies $p(z) = 0$ for $z \leq 0$ then (i) and (ii) alone are sufficient.

The class of admissible penalty functions p is very large. Several practical classes of functions $p(\cdot)$ are discussed below. These are depicted in two dimensions in Figure 10.2. Computationally, a penalty

---

[*] The maximum net improvement is given by

$$\max_{\Delta b} [\overline{v}(b)-\overline{v}(b+\Delta b)-p(\Delta b)] = \overline{v}(b)-V(p) \geq 0 ;$$

by (10.7), the gross improvement $\overline{v}(b)-\overline{v}(b+\hat{\Delta b}) \geq p(\hat{\Delta b})$.

$p(\cdot)$ may best be handled by slightly reformulating (G) to include additional variables and constraints that reflect the nature of $p(\cdot)$.

(A)
$$p(\Delta b) \equiv \sum_{i=1}^{m} \lambda_i \cdot \max \{0, \phi_i(\Delta b_i)\} \qquad (10.9)$$

where $\lambda_i \geq 0$, $(i = 1,\ldots,m)$, $\Delta b \equiv G(x,y)-b$, and

$$\phi_i \equiv \begin{cases} \text{convex function of } \Delta b_i \text{ where } \phi_i(0) = 0, \ \Delta b_i \leq \delta_i \\ M \text{ (very large number), } \Delta b_i > \delta_i . \end{cases}$$

A penalty is incurred only for positive increments Δb; a very large (effectively, an infinite) penalty is assigned to a Δb that exceeds a prescribed upper bound δ. Let $\Phi(\cdot) = (\phi_1(\cdot),\ldots,\phi_n(\cdot))$. This type of penalty $p(\cdot)$ can be implemented by modifying (G) as follows:

$$\begin{array}{l} \text{Minimize } f(x,y) + \lambda \cdot \Phi(z) \, | \, G(x,y)-z \leq b . \qquad (10.10) \\ (x,y) \epsilon U \\ 0 \leq z \leq \delta \end{array}$$

This relaxation of (P) permits what will be referred to as "elastic resources," b+z. Conventional unconstrained optimization techniques might well be an integral part of a solution approach to (G) (provided all functions are sufficiently well-behaved), especially when U reflects only variable bounds and integrality constraints.

(B)
$$p(\Delta b) \equiv \lambda \cdot \Delta b, \text{ where } \Delta b = G(x,y)-b, \ \lambda \geq 0 . \qquad (10.11)$$

This yields the familiar generalized Lagrangean Relaxation of problem (P).[*] Notice that (G) with U replaced by Co(U) -- call this problem (G_*) -- is a relaxation of (P_*). When (P) is linear, a computationally attractive relationship holds between (P), (G_*)

---

[*] Cf. Everett [1], Roode [15], and Geoffrion [5,6]. Only the last of these addresses Lagrangean Relaxation in the context of integer programming, however; other references are cited in [6].

and (P_*): an optimal solution to (G_*) is given by one of the extreme (vertex) points of $Cc(U)$. This solution is feasible in (P) at a perturbed right-side $\hat{b}$, and is therefore optimal at $\hat{b}$ in all three problems, (P), (P_*) and (G_*). Consequently, an alternative optimal solution to (P) may be obtained by solving the LP problem (G_*). This property, of course, is implementable only when $Co(U)$ can be characterized in terms of linear constraints (for instance, when $Co(U) = U_0$ as was discussed earlier).

(C) $p(\Delta b)$ piecewise-linear, possibly discontinuous (see Fig. 10.1c). A fixed charge is incurred for increments Δb_i exceeding a prescribed threshold δ_i. Problem (G) can be expressed as an MIP by introducing integer variables corresponding to the discrete charges and continuous variables corresponding to each linear piece of $p(\cdot)$.
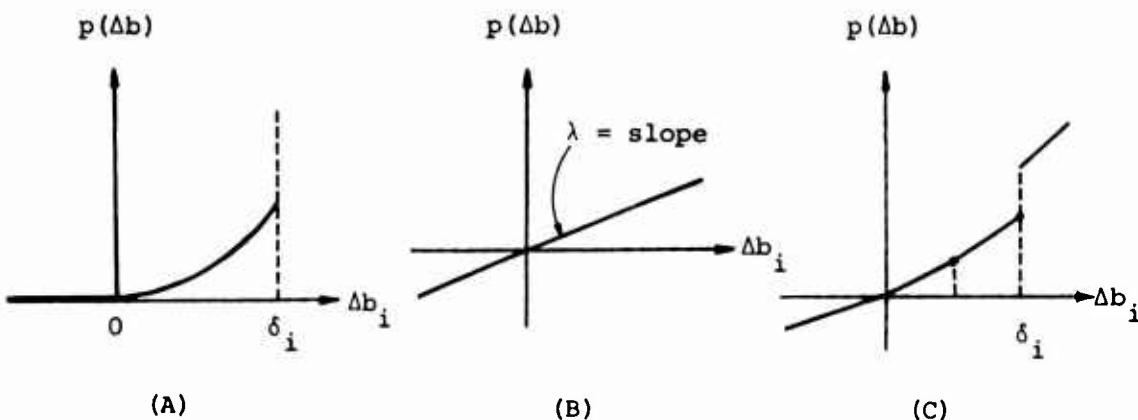
Figure 10.2.   Example Types of Admissible Functions $p(\cdot)$

APPENDIX 10A

APPROXIMATING LOCAL BEHAVIOR OF THE OPTIMAL VALUE FUNCTION

Let y^* be optimal in (P) with RS $\bar{b}$; the subproblem (P_{y^*}) is studied as an indicator of the local behavior of $\bar{v}(b)$. Dual variables μ^* associated with $\bar{b}$, optimal in (P_{y^*}), can be interpreted (with the proper sign convention) as a subgradient of $\bar{v}(\cdot)$ at $\bar{b}$.

Application of the duals μ^* from problem (P_{y^*}) to problem (P) is implied by the following results.

__Theorem.__ If f and G are l.s.c. on X @ Y and if y^* is uniquely optimal at $\bar{b}$, then there exists a $\Delta b > 0$ such that

$$\bar{v}(b) = \bar{v}_{y^*}(b), \quad \text{for} \quad \bar{b} \leq b \leq \bar{b} + \Delta b . \qquad (10A.1)$$

Consequently, μ^* is a subgradient to both $\bar{v}(\cdot)$ and $\bar{v}_{y^*}(\cdot)$ at $\bar{b}$. If, in addition, y^* is strongly feasible at $\bar{b}$, then there exists a $\Delta b' > 0$ such that[*/]

$$\bar{v}(b) = \bar{v}_{y^*}(b), \quad \text{for} \quad \bar{b} - \Delta b' \leq b \leq \bar{b} + \Delta b' . \qquad (10A.2)$$

__Proof:__ The feasible region $F(\cdot)$ is a closed map by Theorem 3.9; by Theorem 3.11 there exists a neighborhood $N(\bar{b})$ such that $y \in Y$ is feasible at $b \in N(\bar{b})$ only if it is feasible at $\bar{b}$. That is, if y is infeasible at $\bar{b}$ then y is infeasible at all $b \in N(\bar{b})$. Moreover, there exists an $\epsilon > 0$ such that for any y feasible at $\bar{b}$, $y \neq y^*$,

$$\bar{v}_y(\bar{b}) \geq \bar{v}(\bar{b}) + \epsilon \qquad (10A.3)$$

since y^* is uniquely optimal and Y is finite. By Theorem 3.10, $\bar{v}_y(\cdot)$ is

---

[*/] The condition $\bar{b} \leq b \leq \bar{b}+\Delta b$ for some $\Delta b > 0$ can be replaced by the more general conditions that $F(b) \supseteq F(\bar{b})$ and $b \in N(\bar{b})$ where $N(b)$ is some neighborhood of $\bar{b}$ and $F(\cdot)$ denotes the feasible region.

l.s.c. at $\bar{b}$, $y \in Y$. Hence, there exists a neighborhood $N_\delta(\bar{b})$, depending on ε, such that for any y feasible at $\bar{b}$,[*]

$$\bar{v}_y(b) < v_y(b) + \varepsilon, \quad b \in N_\delta(\bar{b}) . \qquad (10A.4)$$

Combining (10A.3) and (10A.4), $\bar{v}(\bar{b}) < v_y(b)$. But, for all $b \ni F(b) \supseteq F(\bar{b})$, $\bar{v}_{y*}(b) \le \bar{v}_y(\bar{b}) = \bar{v}(\bar{b})$ so that $y*$ is optimal for all $b \in N_\delta(\bar{b}) \cap N(\bar{b})$ for which $F(b) \supseteq F(\bar{b})$. This establishes result (10A.1). Result (10A.2) follows from Theorem 4.6 ($\exists~N'(\bar{b}) \ni y*$ is optimal at all $b \in N'(\bar{b})$ for which $F(b) \subseteq F(\bar{b})$) and result (10A.1). $\qquad\qquad ||$

In the <u>linear</u> case, (P_{y*}) is an LP problem and $\mu*$ is available from the final tableau of the simplex procedure. Assuming nondegeneracy,[†] there exists a $\hat{b} > \bar{b}$ such that $\bar{v}_{y*}(b) = \bar{v}_{y*}(\bar{b}) + \mu* \cdot (b-b)$ where $\bar{b} \le b \le \hat{b}$. By the theorem, if $y*$ is uniquely optimal in (P),[‡]

$$\Delta\bar{v}(\bar{b}) = \mu*$$

and there exists a $\Delta b > 0$ such that

$$\bar{v}(b) = \bar{v}(\bar{b}) + \mu* \cdot \Delta b, \quad \bar{b} \le b \le \bar{b} + \Delta b . \qquad (10A.5)$$

The weakness in application of the theorem and the shortcoming of dual variable interpretations,[‡] is that Δb is not known and cannot be determined by analyzing subproblem (P_{y*}) alone. This deficiency is depicted in Figure 10A.1 (for one component of b). The size of Δb depends not only on $\bar{v}_{y*}(b)$ but also on all other contending functions,

---

[*] Feasibility of y at $\bar{b}$ is needed for strict inequality in (10A.4).

[†] See Luenberger [10], p. 76.

[‡] Uniqueness is a stronger assumption than is necessary: the conclusions hold as long as $y*$ remains optimal for an increment Δb to b, which can be the case even if $y*$ is not uniquely optimal at $\bar{b}$.

[‡] Reference is to imputation of dual prices; see, for instance, Hespel [7].

$\overline{v}_y(b)$, $y \in Y$.
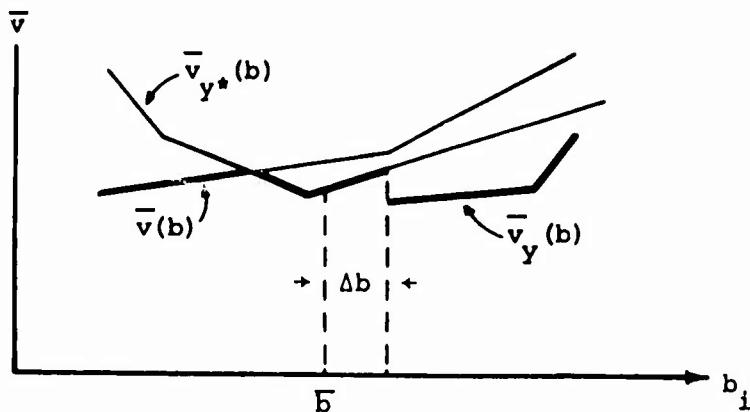
Figure 10A.1. The Size of Δb Depends on $\overline{v}_y$, $y \in Y$.

A ranging analysis that determines or bounds the componentwise

increments to $\overline{b}$ for which (10A.1) or (10A.5) holds is possible by analyz-

ing all subproblem functions $\overline{v}_y$. In the context of branch-and-bound this

amounts to analyzing the subproblem at each fathomed node to determine

or bound the increments Δb_i for which the fathoming conditions remain

valid. Roodman [16,17] exploits this idea and presents an algorithm and

computational results for both pure and mixed integer linear programs.

The major shortcoming of ranging analysis as a means for gaining

sensitivity information is that it has no mechanism for considering

discrete improvements in $\overline{v}$ at points of discontinuity. However, results

may be more significant and meaningful when applied to MIP problems for

which $\overline{v}(b)$ is continuous. Several examples of continuous MIP problems

were mentioned in Part I.[*]

[*]/For example: uncapacitated facility location problems (b = demands);
the nonlinear dynamic lot size model (4.18); the goal programming model
(7.7); capital budgeting problems with both lending and borrowing, such
as (7.5)--see also Weingartner [19], p. 142; another example is the
minimax regression problem studied by Schrage [18].

APPENDIX 10B

A CHARACTERISTIC OF v AND THE ENVELOPE FUNCTION v_*

Let $Q \subseteq B$ denote the set of extreme points of $\overline{v}_*$ (more precisely, of the epigraph of $\overline{v}_*$).

__Theorem__: The functions $\overline{v}(\cdot)$ and $\overline{v}_*(\cdot)$ coincide on Q.

__Proof__: Suppose for some $\hat{b} \in Q$ that $\overline{v}(\hat{b}) = \overline{v}_*(\hat{b}) + \gamma$ where $\gamma > 0$. Let $h(\cdot)$ be any linear support to $\overline{v}_*(\cdot)$ at $\hat{b} \ni h(\hat{b}) = \overline{v}_*(\hat{b})$ and $h(b) < \overline{v}_*(b)$ for $b \neq \hat{b}$ [this is always possible at an extreme point $\hat{b}$]. We will show that

$$\exists\ \delta > 0 \ni \overline{v}(b) > h(b) + \delta, \quad \forall\ b \in B . \tag{10B.1}$$

Given that this property holds the theorem follows easily: the convex hull of the epigraph of $\overline{v}(\cdot)$ is contained in the epigraph of maximum $[h(b), \overline{v}_*(b)]$, which in turn is a proper subset of the epigraph of $\overline{v}_*(\cdot)$ since $\overline{v}_*(\hat{b}) < h(\hat{b}) + \delta$, implying that $\overline{v}_*(\cdot)$ is not the convex envelope of $\overline{v}(\cdot)$. This contradiction then yields the desired result. See Figure 10B.1.

To establish (10B.1) it will first be shown to hold in some neighborhood $N_\sigma(\hat{b})$. Since $\overline{v}(\cdot)$ is l.s.c. $\exists\ \sigma' > 0 \ni \overline{v}(b) > \overline{v}(\hat{b}) - \gamma/2$, $\forall\ b \in N_{\sigma'}(\hat{b})$. Consequently, the assumption that $\overline{v}(\hat{b}) = \overline{v}_*(\hat{b}) + \gamma$ leads to $\overline{v}(b) > \overline{v}_*(\hat{b}) + \gamma/2$ for $b \in N_{\sigma'}(\hat{b})$. Now by continuity of $\overline{v}_*(\cdot)$ on B, $\exists\ \sigma, 0 < \sigma \leq \sigma', \ni \overline{v}_*(\hat{b}) + \gamma/4 > \overline{v}_*(b), \forall\ b \in B \cap N_\sigma(\hat{b})$; therefore,

$$\overline{v}(b) > \overline{v}_*(b) + \gamma/4, \quad \forall\ b \in B \cap N_\sigma(\hat{b}) . \tag{10B.2}$$
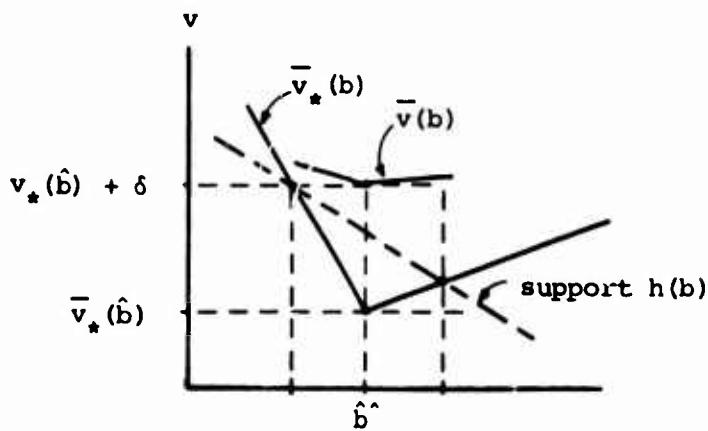
Figure 10B.1. Contradiction: v_* is Not the Convex Envelope if $v(\hat{b}) \neq \bar{v}_*(\hat{b})$

Suppose next that there is no $\delta > 0$ for which (10B.1) holds. Then $\forall n, \exists b \in B \ni \bar{v}(b) < h(b) + \frac{1}{n}$. Thus, $\exists$ a sequence $\{b^{(q)}\} \to \bar{b} \in B$ (B is closed) $\ni \bar{v}(b^{(q)}) \to h(\bar{b})$. Because $\bar{v}_*$ is continuous (uniformly in any closed neighborhood of $\bar{b}$), $\bar{v}(b) \geq \bar{v}_*(b) \geq h(b) \ \forall \ b \Rightarrow \bar{v}_*(\bar{b}^{(q)}) \to \bar{v}_*(\bar{b}) = h(\bar{b}) \Rightarrow \bar{b} = \hat{b}$. But (B2) states that $\bar{v}(\hat{b})$ is bounded away from $\bar{v}_*(\hat{b}) = h(\hat{b})$ in a neighborhood of $\hat{b}$, so that no such sequence $\{b^{(q)}\}$ exists; therefore, $\exists \ \delta > 0 \ni$ (10B.1) holds. ‖

The extreme envelope points are of particular interest in sensitivity analysis since they represent the "low" points on the graph of the optimal value function. The message of the theorem is that at each extreme point $\hat{b}$ there is an optimal solution of the problem (P_*) that is feasible and optimal in the MIP problem (P).

CHAPTER 11

SENSITIVITY TO RIGHT-SIDE DATA: ANALYSIS METHODS

We again address the MIP problem

$$(P) \begin{cases} \text{Minimize } f(x,y) \\ \text{subject to:} \quad (1) \ G(x,y) \leq b \\ \qquad\qquad\quad (2) \ (x,y) \ \varepsilon \ U \end{cases}$$

studied in Chapter 10; a nominal right-side value is denoted by $\bar{b}$.   The

analysis methods discussed in this chapter draw upon the foundational

concepts and properties presented in Chapter 10.

A general approach to RS analysis is described below that relies on

the combined application of several analysis techniques.   Alternative

analysis methods and techniques are briefly examined and the most sig-

nificant of these singled out for a detailed treatment in subsequent

sections.

11.1   A Basic Approach and Alternative Analysis Methods

The analysis proceeds in two phases (cf. §8.2):

(A)   Determine how much improvement in $\bar{v}$ is possible with reason-

able perturbations to the RS data $\bar{b}$.

(B)   Derive alternative solutions that provide substantial net

improvements in $\bar{v}$ with acceptable associated perturbations Δb.

A simple analysis in Phase A may alleviate the need for the more exten-

sive computational requirements in Phase B.   Analysis methods for each

phase will be discussed in turn.

For Phase A there are two computationally inexpensive indicators

of the potential improvement possible in $\bar{v}$.   These are the bounds on

$\bar{v}(b)$ given by (10.1) and (10.3):

$$\bar{v}_W(\bar{b}) + \mu \cdot \Delta b \leq \bar{v}(\bar{b} + \Delta b) \leq \bar{v}_{y*}(\bar{b} + \Delta b) \tag{11.1}$$

where $\bar{v}_W$ and $\mu \leq 0$ correspond to an optimal solution of (P_W) and $\bar{v}_{y*}$ corresponds to (P_{y*}) in which $y*$, optimal in (P) at $\bar{b}$, is held fixed. These bounds can be interpreted as bounds on the amount of improvement $\bar{v}(\bar{b}) - \bar{v}(\bar{b} + \Delta b)$ that is possible. Let δ denote the gap between $\bar{v}$ and $\bar{v}_W$ at $\bar{b}$: $\delta \equiv \bar{v}(\bar{b}) - \bar{v}_W(\bar{b})$. Then for any Δb, the improvement in $\bar{v}(\cdot)$ is bounded as follows (see Figure 11.1):

$$\bar{v}(\bar{b}) - \bar{v}_{y*}(\bar{b} + \Delta b) \leq \bar{v}(\bar{b}) - \bar{v}(\bar{b} + \Delta b) \leq \delta + (-\mu) \cdot \Delta b . \tag{11.2}$$
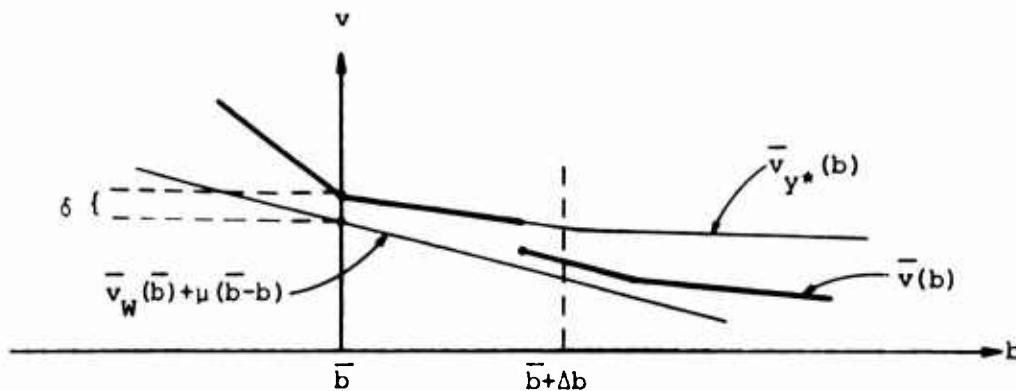
Figure 11.1. The Bounds on $\bar{v}(b)$ Indicate the Potential
for Improvement in $\bar{v}(\bar{b})$

Based on these bounds, two conclusions may possibly be drawn.

(1) The upper bound $\delta + (-\mu) \cdot \Delta b$ is "sufficiently small" for reasonable Δb's that alternative improving solutions are no longer of interest.

(2) The lower bound $\bar{v}(\bar{b}) - \bar{v}_{y*}(\bar{b} + \Delta b)$ is "sufficiently large" (for some Δb) that a search for alternative solutions is worthwhile.

165

If neither of these bounding tests is affirmative then this preliminary

analysis of bounds is inconclusive -- substantial net improvements in $\overline{v}$

may or may not be possible.

Several comments are in order. When the "costs" associated with a

perturbation Δb are nonquantifiable and subjective in nature the bound-

ing tests (1) and (2) are likewise subjective. By "sufficiently small"

is meant an insignificant net improvement in $\overline{v}$ when all types of costs

assigned to Δb are considered. The bound $\delta + (-\mu) \cdot \Delta b$ on the maximum

possible improvement (given Δb) must be weighed against the consequential

costs of the perturbation Δb; although the bound may in some sense be

large, a counteracting penalty on Δb may produce an insignificant net

potential improvement in $\overline{v}$. A similar consideration applies to the term

"sufficiently large" in test (2). These tests can be made more mathe-

matically precise if the major consequential costs can be quantified in

terms of Δb.

We also observe that the analysis required in (1) and (2) is sim-

plified if one considers only directional changes $\overline{b} + \theta \overline{r}$, $\theta \geq 0$, to the

RS $\overline{b}$. Also, in this case an easy-to-compute linear upper bound on $\hat{v}(\theta)$

can be used in place of the function $\overline{v}_{y*}(\overline{b} + \Delta b)$ in (11.1) and (11.2) when

(P) is a convex program: let $y*$ be optimal at $\overline{b}$ ($\theta = 0$) in (P) and let

$\overline{\theta}$ ($\overline{\theta} \neq 0$) be such that $y*$ is feasible in (P) at $\overline{b} + \overline{\theta} \overline{r}$ (e.g., $\overline{\theta}$ can be

any number ≥ 0 if $\overline{r} \geq 0$). Then for all θ, $0 \leq \theta \leq \overline{\theta}$,

$$\hat{v}(\theta) \leq \frac{1}{\overline{\theta}} [(\overline{\theta} - \theta) \hat{v}(0) + \theta \hat{v}_{y*}(\overline{\theta})] \ .$$

This bound derives easily from convexity of $\overline{v}_{y*}$.

In the event that the bounding tests (1) and (2) are inconclusive

one may either proceed to Phase B or make a further attempt to estimate

the sensitivity of $\bar{v}$ to b. In the vein of the latter, the following exploratory run may be made: select a RS $b' \geq \bar{b}$ and solve (P) at b'. By monotonicity, $\bar{v}(\bar{b}) \geq \bar{v}(b) \geq \bar{v}(b')$ for all b such that $\bar{b} \leq b \leq b'$. This leads to two more tests:

(1') The improvement $\bar{v}(\bar{b}) - \bar{v}(b')$ is so small that alternative solutions resulting from perturbed values b $(\bar{b} \leq b \leq b')$ are no longer of interest.

(2') The improvement $\bar{v}(\bar{b}) - \bar{v}(b')$ less the costs assigned to $\Delta b^* \equiv (b'-s^*)-\bar{b}$ (where s^* denotes optimal slack) is substantial -- Phase B is indicated.

These tests also can be inconclusive and one must be careful in drawing conclusions: there can exist perturbations Δb that are much preferred over Δb^* and which yield an improvement comparable in magnitude to $\bar{v}(\bar{b}) - \bar{v}(b')$. If $p(\Delta b^*)$ represents a cost assigned to Δb^* then the quantity $\bar{v}(\bar{b}) - \bar{v}(b') - p(\Delta b^*)$ is a lower bound on the net improvement possible in $\bar{v}(\cdot)$; an encouragingly high value of this lower bound leads to conclusion (2'). Conclusion (1'), however, is based only on the gross difference $\bar{v}(\bar{b}) - \bar{v}(b')$, which represents an upper bound on the improvement arising from a Δb, $0 \leq \Delta b \leq b'-\bar{b}$.

It is worth noting that the upper bound in (11.2) can result as a byproduct of solving (P). [*] By executing an exploratory run an additional bound is generated, thus strengthening the bounding tests.

When the majority of the costs incurred by perturbations Δb can be expressed functionally, relaxation (G) becomes an attractive alternative to the Phase A techniques just discussed. Choose $p(\Delta b)$ as any convenient

---

[*] E.g., an LP start in branch-bound yields the linear lower bound on $\bar{v}$ (upper bound in (11.2)), where $W = U_0$.

conservatively low approximation to the true costs associated with Δb, where $p(\Delta b) = 0$ for $\Delta b \leq 0$ (i.e., a type A penalty; cf. §10.4). Then solving problem (G) leads to a generalization of tests (1') and (2'): if there do not exist alternative solutions that provide a net improvement in $\overline{v}$, then the solution to (G) will satisfy the optimality conditions of Theorem 10.4a. Thus, problem (G) produces in this case both an optimal solution to (P) and proof of the fact that Phase B would be fruitless. On the other hand, a nonzero optimal value of Δb in (G) suggests a Phase B analysis.

Evidently, relaxations of type (G) can also be effective in meeting the Phase B objective: (G) produces an alternative solution to (P) that nets the maximum improvement in $\overline{v}$ based on the chosen criterion $p(\cdot)$. In effect, a solution to (G) takes advantage of discontinuities in $\overline{v}$. Subjective cost considerations can be accommodated by varying the costs $p(\cdot)$ to yield a collection of alternative improving solutions. Ideally, these variations in p would represent the range of both quantifiable and nonquantifiable costs associated with perturbations in the resource availabilities $\overline{b}$. Relaxations of type (G), particularly the elastic resource model (10.10), are exploited further in a later section.

Relaxations built on convexification in domain space (e.g., dropping integrality constraints) generally produce fractional values for the discrete variables y. "Rounding" refers to the rounding up or down of these fractional variables in an effort to obtain a solution that is feasible in (P), and is a natural approach for constructing a solution to pure integer problems. Hopefully, clever rounding of a noninteger solution will produce a near-optimal solution; when this works, the integer solution is obtained at a fraction of the computational cost

168

that would be required by an integer programming algorithm.

The objective for studying rounding here is to consider it not as a means for solving an integer problem, but rather as a possible means for obtaining alternative improving solutions. In particular, we search for situations in which rounding can produce a solution that is feasible and optimal at a perturbed right-side data point. When a rounded solution is not known to be optimal at the associated RS b, then a limit (ϵ) on suboptimality is provided by Theorem 10.2.

Rounding is discussed in detail in the next section. The primary incentive for using rounding to obtain alternative solutions is, of course, the relatively very low computational cost. A shortcoming is the little amount of control over the magnitude of the induced perturbation Δb.

The execution of exploratory runs using alternative right-sides b might also be considered as a Phase B method. To be effective, this method would require solving (P) at several or many RS points b, possibly calling for an application of parametric integer programming (cf. Nauss [12]). The exploration method is straightforward to execute, requires no alterations to the model (except for changes in RS data), and provides direct control over the perturbations Δb. On the other hand, its brute force application can be wasteful in that many undesirable alternative solutions may be generated; even worse, it does not dependably seek out discontinuities in $\bar{v}(\cdot)$ and the most preferred alternative solutions can remain undiscovered.

Restricting the RS data to vary in parametric fashion, $\bar{b}+\theta\bar{r}$, tends to make the analysis problem more comprehensible. The optimal value $\hat{v}$ is a function of the scalar variable θ, the bounding tests are simplified,

169

and the relationships between (P) and its relaxations per Theorem 10.4
(and §11.3) are easily specialized. Although obvious simplifications
result by considering parametric RS, the suitability of such a restric-
tion for the particular application must be considered -- in some in-
stances a parametric analysis can provide only partial sensitivity
information and cannot dependably produce the most preferred alterna-
tive solutions.

11.2 Rounding

A special rounding property of knapsack problems[*] motivates in-
vestigation of rounding as a means for sensitivity analysis in more
general applications. The knapsack problem is stated as follows:

$$
\begin{cases}
\text{Maximize} \quad \sum\limits_{i=1}^{n} c_i y_i \\
0 \leq y_i \leq k_i \\
\quad \text{and} \\
\text{integer} \\
\text{subject to:} \quad \sum\limits_{i=1}^{n} a_i y_i \leq b
\end{cases}
\tag{11.3}
$$

where it is assumed that $c \geq 0$ and $a \geq 0$. It is well-known that an
optimal solution to the LP relaxation (i.e., the LP problem obtained by
dropping the integrality requirements) may have at most one fractional
variable and that rounding this fractional variable either up or down
yields a solution that is optimal in (11.3) for some modified right-side
value b'.

To be specific, let y^0 denote an optimal solution to the LP relax-
ation of (11.3). If y^0 is integral in each component then it is obvi-
ously feasible and optimal in (11.3). If y^0 is not integral in each

---

[*] Cf. Garfinkel and Nemhauser [3], Chapter 6.

170

component then exactly one component is fractional, say y_q^0. Let $\underline{y}$

differ from y^0 only in that the fractional component is replaced by its

integer part, $[y_q^0]$; similarly, let $\overline{y}$ differ from y^0 only in that the

fractional component is replaced by its integer part plus one: $[y_q^0]+1$.

Secondly, define $\underline{b} \equiv \Sigma a_i \underline{y}_i$ and $\overline{b} \equiv \Sigma a_i \overline{y}_i$. Then $\underline{y}$ is optimal at $\underline{b}$ and $\overline{y}$

is optimal at $\overline{b}$ in (11.3).

The optimal value function $\overline{v}_0(b)$ for the LP relaxation of (11.3)

is concave and piecewise linear on $(0,\infty)$ with n' break points ($n' \leq n$).

It is interesting to note that each of these n' values of b (and only

these values) yields a natural integer solution to (11.3). This sug-

gests that naturally integer solutions might be found to more general

MIP problems by identifying break points of the continuous envelope

function $\overline{v}_*(\cdot)$ [in problem (11.3), $\overline{v}_0(\cdot) = \overline{v}_*(\cdot)$].

With reference again to problem (P), a rather obvious procedure

for using rounding to obtain an alternate solution is suggested by

Theorem 10.2. Select a convex set $W \supset U$; assume that f and G are con-

vex on W:

(a) Solve relaxation (P_W) at $\overline{b}$, obtaining $(\overline{x},\overline{y})$ and an optimal

multiplier vector $\mu \leq 0$ (assuming one exists).

(b) If $\overline{y}$ is fractional, round $\overline{y}$ to $\hat{y}$ (and adjust $\overline{x}$ if necessary

to some value $\hat{x}$) so that $(\hat{x},\hat{y})$ is feasible in (P) except for

possible violation of constraints (1).

(c) Adjust $\overline{b}$ to b' if necessary so that $(\hat{x},\hat{y})$ is feasible at b'

in (P).

(d) Then $(\hat{x},\hat{y})$ is ϵ-optimal in (P) at the perturbed RS b', where

$\epsilon = f(\hat{x},\hat{y}) - f(\overline{x},\overline{y}) - \mu(b'-\overline{b})$.

171

Of course, whenever possible, steps (b) and (c) should be executed in a way that tends to minimize ε. If Co(U) can be characterized in terms of a set of convex constraints and is selected as the set W, then Corollary 10.3.1 assures that there exists a $(\hat{x},\hat{y})$ such that $\varepsilon = 0$; but this does not imply that such a solution can be generated via rounding of $\bar{y}$.

We will show that for the following class of problems an "optimal rounding" -- that is, a rounding that is optimal at some perturbed RS point -- is always possible; in fact, when the set U involves only bounds and integrality constraints on the y variables, then an arbitrary rounding of $\bar{y}$ is an optimal rounding, where $\bar{y}$ is optimal in the relaxation obtained by simply dropping integrality restrictions.

$$(SP) \begin{cases} \text{Minimize } f(x) + cy \\ \text{subject to: } (1)\ G(x) + Ay \leq \bar{b} \\ \qquad\qquad\quad (2)\ x \in X \\ \qquad\qquad\quad (3)\ y \in Y \end{cases}$$

where c and A are constants of the appropriate dimensions and Co(Y) is polyhedral; no assumptions on f or G are needed. As in previous notation, Y in constraints (3) replaced by Y_0 yields the problem (SP_0) and Y replaced by Co(Y) yields (SP_*) since $Co(X \times Y) = X \times Co(Y)$.

Theorem 11.1: (a) Let $(\bar{x},\bar{y})$ and $\bar{\mu}$ be optimal in (SP_*) where $\bar{\mu}$ is a multiplier vector associated with $\bar{b}$. There exist roundings $\hat{y}$ of $\bar{y}$ such that $(\bar{x},\hat{y})$ is optimal in (SP) with $\bar{b}$ perturbed to any b' satisfying[*]

---

[*] The conditions $\bar{\mu}_i = 0$, $\bar{\mu}_i < 0$ in (11.4) can be replaced by the slightly stronger ones: $\bar{s}_i > 0$, $\bar{s}_i = 0$, respectively, where $\bar{s}$ denotes the slack $G(\bar{x})+A\bar{y}-\bar{b}$.

172

$$b_i' \begin{cases} \geq [G(\overline{x}) + A\hat{y}]_i, & \text{if } \overline{\mu}_i = 0 \\ = [G(\overline{x}) + A\hat{y}]_i, & \text{if } \overline{\mu}_i < 0 . \end{cases} \tag{11.4}$$

(b) Suppose that $Y = \{y \mid 0 \leq y \leq u, y \text{ integer}\}$ where the bound u is integer. Let $(\overline{x}, \overline{y})$ and $\overline{\mu}$ be optimal in (SP_0). Arbitrarily round $\overline{y}$ to $\hat{y}$ (or replace $\overline{y}$ by any $\hat{y} \in Y$). Then $(\overline{x}, \hat{y})$ is optimal in (SP) at any b' satisfying (11.4).

Proof: Project (SP) onto the x variables:

$$\underset{x \in X}{\text{Min}} [f(x) + \underset{y \in Y}{\text{min}} cy \mid Ay \leq \overline{b} - G(x)] . \tag{11.5}$$

A Lagrangean Relaxation of the inner problem is given by

$$\underset{y \in Y}{\text{Min}} \ cy + \overline{\mu}[\overline{b} - G(\overline{x}) - Ay]$$

which can be rewritten as the LP problem

$$\overline{\mu}[\overline{b} - G(\overline{x})] + \underset{y \in Co(Y)}{\text{Min}} (c - \overline{\mu}A)y . \tag{11.6}$$

Now $\overline{y}$ and $\overline{\mu}$ must satisfy the conditions:

(i) $\overline{y}$ minimizes $(c - \overline{\mu}A)y$ s.t. $y \in Co(Y)$,

(ii) $A\overline{y} \leq \overline{b} - G(\overline{x})$,

(iii) $\overline{\mu}[\overline{b} - G(\overline{x}) - A\overline{y}] = 0$.

Result (a) can be reasoned as follows: if $\overline{y}$ is fractional then it is not an extreme point of Co(Y) -- it must lie on an edge or face of Co(Y) and every point on this edge (or face) of Co(Y) is optimal in (11.6); this includes some roundings of $\overline{y}$ (e.g., the extreme points of the face).

In part (b), $Co(Y) = Y_0$ $\{y \mid 0 \leq y \leq u\}$. Notice that if $\overline{y}_j$ is fractional then $(c - \overline{\mu}A)_j = 0$ since otherwise $\overline{y}_j$ would not be minimizing in (11.6) subject to $0 \leq y_j \leq u_j$. Replace the fractional $\overline{y}_j$ by any $\hat{y}_j$ such

that $\hat{y} \in Y$; let b' satisfy (11.4). Then by construction, $(\overline{x}, \hat{y})$ and $\overline{\mu}$ satisfy the optimality conditions:

(i) $(\overline{x}, \hat{y})$ minimizes $f(x) + cy + \overline{\mu}[b' - G(x) - Ay]$ subject to

$x \in X$, $y \in Y_0$ (since $Co(X \times Y) = X \times Y_0$)

(ii) $G(\overline{x}) + A\hat{y} \leq b'$

(iii) $\overline{\mu}(b' - G(\overline{x}) - A\hat{y}) = 0$

and therefore $(\overline{x}, \hat{y})$ must be optimal in (SP). ||

The proof reveals the utility of Lagrangean Relaxation for identifying (simultaneous) post-optimality ranges on the RS b.

In case (a) of the theorem roundings $\hat{y}$ must be selected so that they satisfy all constraints represented by the set Y. Not all such feasible roundings are optimal in (SP) for some RS b. The special structure of Y in part (b) assures that any rounding is feasible in (SP) for some b'. For more complex sets Y, a rounding of a $\overline{y}$ optimal in (SP_0) might not be optimal at any perturbed RS b'; moreover, it is possible that no such optimal roundings exist, as is demonstrated by the following example.

$$\underset{y \in Y}{\text{Maximize}} \ 20y_1 + 22y_2 + y_3 \ | \ 4y_1 + y_2 \leq z = \overline{b} \qquad (11.7)$$

where $Y = \{y | y_1, y_2, y_3 = 0 \text{ or } 1; \ y_1 + 2y_2 \leq 2; \ 14y_1 + 7y_2 + y_3 \leq 13\}$. An optimal solution to the LP relaxation obtained by dropping integrality is given by $\overline{y} = (4/7, 5/7, 0)$. Only one rounding of $\overline{y}$, other than the trivial one $(0,0,0)$, is contained in Y: $\dot{y} = (0,1,0)$. Correspondingly, b' = 1. But $\dot{y}$ is not optimal in (11.7) at b' -- $y = (0,1,1)$ is uniquely optimal (and this is not a rounding of $\overline{y}$).

The following example, examined by Rappaport [14] and by Jensen [8], illustrates the simplicity and practicality of using rounding to

174

obtain alternative (optimal) solutions.

$$\text{Maximize}_{y \in Y} \quad 25y_1 + 30y_2 + 40y_3 + 100y_4$$

$$\text{subject to:} \quad 12y_1 + 15y_2 + 15y_3 + 25y_4 \leq 3500 \qquad (11.8)$$

$$10y_1 + 12y_2 + 20y_3 + 20y_4 \leq 3000$$

$$0.5y_1 + 0.6y_2 + 0.6y_3 + 2y_4 \leq 240$$

where $Y = \{y \,|\, y \text{ integer}, y_1 \geq 10, y_2 \geq 10\}$.

This problem satisfies the conditions of Theorem 11.1(b); hence, any rounding of an optimal LP solution is optimal in (11.8) at an appropriately perturbed RS b. Rappaport obtained an "acceptable" solution to (11.8) by solving the LP relaxation and rounding down the fractional variables (thus assuring a feasible solution). Jensen shows in [8] that a better integer solution exists and presents the following computational results:

LP Relaxation		Rounded Solution	Optimal Solution
$\bar{v}_0 = 12{,}490$		$= 12{,}410$	$\bar{v} = 12{,}470$
$\bar{y}_1 = 10$	$\bar{\mu}_1 = 4/3$	$y_1 = 10$	$y_1 = 10$
$\bar{y}_2 = 10$	$\bar{\mu}_2 = 0$	$y_2 = 10$	$y_2 = 10$
$\bar{y}_3 = 49$	$= 100/3$	$y_3 = 49$	$y_3 = 48$
$\bar{y}_4 = 99.8$		$y_4 = 99$	$y_4 = 100$

If we now abandon the assumption of inelastic resource availabilities implicit in the above results, then a conspicuous integer solution is the one obtained by rounding y_4 to 100. This requires $\Delta b = (5,0,0.4)$

and yields an improved objective value of 12,510. This rounded solution is optimal by Theorem 11.1(b); as a check, $\epsilon = c \cdot \Delta y - \overline{\mu} \cdot \Delta b = 20 - [(4/3)(5) + (100/3)(0.4)] = 0$. Notice that the optimality range on b' allows $b_2' \geq 2857$ ($\overline{\mu}_2 = 0$).

In practice one would want to round $\overline{y}$ in such a way that the induced perturbation to b is acceptable. For large problems this would require a systematic approach that can be automated. There are many ways that integer problems could be formulated to derive roundings. The following result considers one possible choice. Let $L \equiv \{j \mid \overline{y}_j$ fractional in $(SP_0)\}$ and suppose that $\delta b \geq 0$ is a limit on the permissible perturbation Δb. Then an acceptable rounding (of the fractional components of $\overline{y}$) is given by the following 0-1 integer problem:

$$\begin{cases} \text{Minimize} \quad \sum_{j \in L} c_j y_j \\ y = 0,1 \\ \text{subject to:} \quad \hat{A}y \leq \overline{b} + \delta b - A[\overline{y}] \end{cases} \tag{11.9}$$

where $\hat{A}$ denotes A with columns j, j $\notin$ L, removed; $[\overline{y}]$ denotes $\overline{y}$ with all fractional components rounded down to the nearest integer. The next result is immediate from Theorem 11.1(b).

Corollary 11.1.1: Assume that Y is of the form hypothesized in Theorem 11.1(b). If y* is optimal in (11.9) then $\hat{y} \equiv [\overline{y}]+y^*$ is optimal in (SP) for all right-sides b' satisfying (11.4).

Problem (11.9) produces a rounding of $\overline{y}$ that makes the best use of the available resource reserves. By the corollary, $[\overline{y}]+y^*$ is an optimal alternative solution to (SP) (at the perturbed point $\hat{b}$). When the number of fractional variables is small relative to the total number n of variables, this alternative solution is obtained very cheaply -- an

176

LP version of the MIP plus a small 0-1 integer problem. Since the number of fractional variables cannot exceed the number m of constraints one is assured that (11.9) is small relative to (SP) if m is small relative to n.

Computational experience indicates that the number of fractional variables arising in LP resource allocation problems is generally considerably less than the number of constraints m. Even when n and m are of comparable size one can expect that problem (11.9) will be a relatively small problem compared to (SP).

As an alternative to the rounding problem (11.9) one could select as an objective the minimization of ϵ given by Theorem 10.2. This then would produce an optimal rounding if one exists and applies to problems more general than those permitted by Theorem 11.1(b). As an illustration of this technique and application of the preceding theorems, consider the production scheduling problem (with I activities, T periods, and a single resource) discussed by Lasdon [9]:

$$\text{Minimize} \sum_{i,j} c_{ij}\theta_{ij}$$

$$\text{subject to:} \sum_{i,j} a_{ij}^{t}\theta_{ij} \leq b_t, \quad (t = 1,\ldots,T)$$

$$Y: \begin{cases} \sum_{j} \theta_{ij} = 1, & (i = 1,\ldots,I) \\[2mm] \theta_{ij} \geq 0, & (i = 1,\ldots,I;\ j = 1,\ldots,N_i) \\[2mm] \theta_{ij} = \text{integer} . \end{cases} \quad (11.10)$$

This problem is of the form (SP) and since Y involves only bounds and GUB constraints on θ, $\text{Co}(Y) = Y_0$. Hence, by Theorem 11.1(a), there exists an optimal rounding of $\bar{\theta}$, optimal in the LP relaxation. Notice

177

that if $I >> T$ then most of the components of $\overline{\theta}$ are naturally integer (for at least $I-T$ indices i, $\overline{\theta}_{ij} = 1$ for some j and 0 for all other j). If the number of fractional variables is small then any rounding that satisfies the GUB constraints should not appreciably violate the resource constraints.

Let $\overline{\theta}, \overline{\mu}$ be optimal primal and dual variables at $\overline{b}$ in the LP relaxation of (11.10). Using any criteria whatever, round $\overline{\theta}$ (to $\hat{\theta}$) in such a way that the GUB constraints ($\Sigma \hat{\theta}_{ij} = 1$) are satisfied. Let $\hat{b}$ be such that $\hat{\theta}$ satisfies the resource constraints at $\hat{b}$. Then $\hat{\theta}$ is ε-optimal at $\hat{b}$ where

$$\varepsilon \equiv \sum_{i,j} c_{ij}(\hat{\theta}_{ij} - \overline{\theta}_{ij}) - \sum_{k=1}^{T} \overline{\mu}_k(\hat{b}_k - b_k) . \qquad (11.11)$$

The problem of finding a rounding $\hat{\theta}$ that is optimal at some point $\hat{b}$ in (11.10) can be expressed as a mixed integer problem (which is much smaller in size than the original). Stated in words, this problem is to select $\Delta\theta$ and Δb to minimize $c \cdot \Delta\theta - \overline{\mu} \cdot \Delta b$ (i.e., ε, given by (11.11)) where $\overline{\theta} + \Delta\theta$ is feasible at $b + \Delta b$ in (11.10) and $\Delta\theta_{ij} = 0$ if $\overline{\theta}_{ij} = 0$ or 1. This objective quantity is nonnegative and when equal to zero, $\theta + \Delta\theta$ is optimal at $b + \Delta b$.

To express this problem mathematically we will need the following index sets:

$$L \equiv \{i | \overline{\theta}_{ij} \text{ is fractional for some } j\}$$
$$L_i \equiv \{j | \overline{\theta}_{ij} \text{ is fractional}\} .$$

The 0-1 problem is to

$$\text{Minimize} \quad \sum_{i \in L} \sum_{j \in L_i} c_{ij}(y_{ij} - \bar{\theta}_{ij}) - \sum_{t=1}^{T} \bar{\mu}_t z_t$$
$$_{y,z}$$

$$\text{subject to:} \quad \sum_{i \in L} \sum_{j \in L_i} a_{ij}^t(y_{ij} - \bar{\theta}_{ij}) - z_t = b_t - \sum_{i,j} a_{ij}^t \bar{\theta}_{ij} \qquad (11.12)$$

$$\sum_{j \in L_i} y_{ij} = 1, \quad (i \in L)$$

$$y_{ij} = 0 \text{ or } 1, \quad (i \in L, \, j \in L_i) \; .$$

In this formulation Δb is represented by the z variable and $\Delta \theta$ by $y_{ij} - \bar{\theta}_{ij}$ for $i \in L$ and $j \in L_i$ ($\Delta \theta_{ij} = 0$ for all other i,j).

Geometrically, problem (11.12) accomplishes finding a point $\hat{b}$ on the graph of $\bar{v}(b)$ that minimizes the difference between $\bar{v}(\hat{b})$ and the value of the linear function $\bar{v}_0(b) + \bar{\mu}(\hat{b} - b)$. For problem (11.12) this difference is zero. In general, alternative optimal $\hat{b}$ exist.

In summary, it has been shown that for general bounded linear MIP problems an easily calculated quantity provides an upper bound on the degree of suboptimality of a rounded solution. Problems of the form (SP) have the property that a rounding of $\bar{y}$, optimal in (SP$_*$), exists which is optimal in (SP) at some perturbed RS b'; when Y involves only bounds and integrality constraints on y, every rounding of $\bar{y}$, optimal in (SP$_0$), is optimal in (SP) at some RS b'. An optimal or near-optimal (at a perturbed point b') rounding of the LP solution can be realized by solving an IP problem that is much smaller in size than the original.

Probably the greatest shortcoming of a rounding approach is that there is little control over the perturbation Δb, except for the special problems of type (SP) which permit control of Δb per Corollary 11.1.1. The only control over Δb is via the selection of a particular rounding. If an "optimal" or near-optimal rounding (at $\hat{b}$) requires a Δb that is

179

unacceptable, then a search for an alternative rounding requiring a more reasonable perturbation of b amounts to a tradeoff between sub-optimality and the magnitude of Δb.

On the plus side, there are many applications in which rounding can produce desirable results. A particularly significant incentive for using rounding is the relatively very low computational cost for obtaining information.

11.3 Elastic Resource Models

Use of relaxations of the type (G) studied in Chapter 10 and elastic resource models such as (10.10) as means for deriving alternative solutions to (P) has already been emphasized.

The purpose of this section is to first give a brief discussion of the properties and use of elastic resource (ER) models for right-side analysis, and second, to present additional properties that result when a parametric penalty is used.

The special case of relaxation (G) that gives zero co.. to over-supplied resources is the elastic resource problem:

$$
(Q) \quad \begin{cases}
\text{Minimize } f(\overline{c};x,y) + p(z) \\
\quad x,y,z \\
\\
\text{subject to: } G(\overline{a};x,y) - z \le \overline{b} \\
\\
\quad\quad\quad\quad (x,y) \ \varepsilon \ U \\
\\
\quad\quad\quad\quad z \ge 0
\end{cases}
$$

where it is assumed that $p(0) = 0$. An optimal solution is denoted by $(\hat{x},\hat{y};\hat{z})$. Equality resource constraints can be accommodated by expressing them as inequalities.

The optimal value function $\overline{V}(p)$ is continuous with respect to p,

i.e., the data defining p, as well as with respect to a, b, and c, provided that p itself is continuous with respect to the data (this does not preclude discontinuity of p with respect to the variables z). These favorable continuity properties make this function more attractive to analyze than the generally discontinuous function $\bar{v}(b)$. The fundamental relationships between $\bar{V}(p)$ and $\bar{v}(b)$ are furnished by Theorem 10.4.

This relaxation of (I) obtained by introducing the continuous non-negative variables z gives recourse to the resource constraints by allowing them to be violated for a price. Of all nonnegative perturbations to $\bar{b}$ in (P), $\hat{z}$ is preferred in the sense that it affords the greatest net improvement in optimal value, based on the criterion $p(\cdot)$.

When it happens that $\hat{z} = 0$, then additional resources are more costly than they are worth. In this case the optimal solution obtained for (Q) is feasible and optimal at $\bar{b}$ in (P). This situation also provides a bound on the sensitivity of $\bar{v}$ to perturbations in $\bar{b}$: a perturbation Δb cannot give rise to an improvement larger than $p(\Delta b)$ (cf. Th. 10.4c); that is,

$$\bar{v}(\bar{b}) - \bar{v}(\bar{b}+\Delta b) \le p(\Delta b) \quad \text{for all} \quad \Delta b \in E^m . \tag{11.13}$$

The perturbation $\hat{z}$ produced by (Q) is never larger than is necessary to accommodate the obtained solution $(\hat{x}, \hat{y})$. This is expressed as a complementary slackness property:

Proposition 11.2: Let $\hat{s}$ denote optimal slack in (Q). Then $(\hat{x}, \hat{y}, \hat{s})$ is optimal at $\bar{b}+\hat{z}$ in (P) and $\hat{s} \cdot \hat{z} = 0$ (provided that $p(\cdot)$ is monotone increasing).

Proof: Extend the definition of $p(\cdot)$ to all of E^m thusly:

181

$$p(z_1,\ldots,z_i,\ldots,z_m) = p(z_1,\ldots,0,\ldots,z_m), \quad \text{if} \quad z_i < 0 .$$

Then $p(\hat{z}) \le p(\hat{z}-\hat{s})$ since

$$\overline{v}(p) = p(\hat{z}) + \overline{v}(\overline{b}+\hat{z}) \qquad \text{[by Th. 10.4c]}$$

$$\le p(\hat{z}-\hat{s}) + \overline{v}(\overline{b}+\hat{z}-\hat{s}) \qquad \text{[by Th. 10.4d]}$$

$$= p(\hat{z}-\hat{s}) + \overline{v}(\overline{b}+\hat{z}) \qquad \text{[by optimality of } \hat{s}] .$$

Now if $\hat{z}_i > 0$ for some i (i = 1,...,m), then $\hat{s}_i = 0$ by (strict) monoto-
nicity of p for nonnegative arguments -- (if $\hat{s}_i > 0$, then $\hat{z}_j \ge$
$\max\{0, \hat{z}_j-\hat{s}_j\}$ for all j and $\hat{z}_i > \max\{0, \hat{z}_i-\hat{s}_i\}$ so that $p(\hat{z}) > p(\hat{z}-\hat{s})$,
a contradiction). $\qquad\qquad ||$

Typically, $\hat{z} \ne 0$ implies that $\overline{b}+\hat{z}$ is a point of discontinuity of
$\overline{v}(\cdot)$; in fact, this is always the case for pure integer problems·

Proposition 11.3: Suppose that (P) is a pure integer problem. If
$\hat{z} \ne 0$ where $\hat{z}$ is optimal in (Q) for any p, then $\overline{v}(\cdot)$ is discontinuous
at $b = \overline{b}+\hat{z}$.

Proof: The result follows from the characterization of $\overline{v}(\cdot)$ as a finite
set of constants: define $R \equiv \{b \in E^m | \overline{v}(b) < \infty\}$ and $\overline{W}_y \equiv \{b \in R | \overline{v}(b) = \overline{v}_y(b)\}$, $y \in Y$. That is, $\overline{W}_y$ is the subset of R on which y is optimal and
$\overline{v}(b) = f(\overline{c};y)$ for all $b \in \overline{W}_y$. Since $R = \bigcup_{y \in Y} \overline{W}_y$ and Y is a finite set,
it follows that R is covered by a finite number of sets[*/] on each of
which $\overline{v}(\cdot)$ is constant.

Suppose now that $\overline{v}(\cdot)$ is continuous at $\overline{b}+\hat{z}$. Then $\exists$ a b' < $\overline{b}+\hat{z}$ ∋
$\overline{v}(b') = \overline{v}(\overline{b}+\hat{z})$, where equality holds per the above characterization.

[*/] Incidentally, it is straightforward to show that the "optimality"
parameter sets $\overline{W}_y$ are polyhedral if all constraints are linear.

But then complementary slackness is violated since $\hat{s} \geq \bar{b}+\hat{z}-b' > 0$ and $\hat{z} \nleq 0$. $\quad\quad\quad\quad ||$

Problem (Q) does not place explicit bounds on the perturbation z to $\bar{b}$; however, such bounds may be enforced effectively via high penalties for large values of z. For instance, requiring that $z \in S \equiv \{z \in E^m | 0 \leq z \leq u\}$ is equivalent to defining $p(z) = \infty$ for $z \notin S$. Since Theorem 10.4 applies to discontinuous functions $p(\cdot)$, it remains valid when the bounds $z \in S$ are added to problem (Q). This alleviates the need to define $p(\cdot)$ outside of the set S and avoids having to deal with too complex a function $p(\cdot)$.

Because the main goal is to derive alternative improving points and solutions, the accuracy with which $p(\cdot)$ represents the decision-maker's real costs is not critical -- the penalty function is primarily a device for discovering "reasonable" perturbations in b that give rise to discrete or relatively large improvements in optimal value. A linear parametric representation of $p(\cdot)$ will be sufficient for many applications.

Let θ be a scalar parameter. A parametric penalty linear in θ is given by $\theta q(z)+r(z)$. When this penalty is substituted for $p(z)$ in (Q) the optimal value can be viewed as a function of the single variable θ. If $r(\cdot)$ is chosen as a lower bound on the decision-maker's true penalty and $r(\cdot)+q(\cdot)$ as an upper bound, then the pertinent values of θ lie in the unit interval $[0,1]$. Since θ appears only in the objective function, the theory and analysis methods of Chapter 9 apply.

Use of the elastic resource model (Q) with a parametric penalty in effect converts the problem of right-side analysis to a problem parametric in the objective function. Deriving $V(\theta)$ for values of θ between

183

0 and 1 produces a set of alternative solutions to (P). The function $V(\theta)$ is concave and monotone nondecreasing with respect to θ; this holds for arbitrary functions f, G, q and r. If the parametric penalty is taken as $\theta q(z)$, then these properties imply that for any optimal solution $(\hat{x}(\theta),\hat{y}(\theta);\hat{z}(\theta))$, the improved optimal value $\overline{v}(\overline{b}+\hat{z}(\theta)) = f(\hat{x}(\theta),\hat{y}(\theta))$ is monotone nondecreasing and the quantity $q(\hat{z}(\theta))$, the slope of $V(\cdot)$ at θ, is nonincreasing.

Alternative solutions to the following capital budgeting problem were generated by deriving $V(\theta)$ using the procedure described in Appendix 9A.

$$\overline{v}(b) = \underset{\substack{y=0,1}}{\text{Max}} \ \overset{8}{\underset{i=1}{\Sigma}} \ c_i y_i \ \Big| \ Ay \leq b \tag{11.14}$$

where $\quad c \equiv (35, 85, 135, 27, 94, 10, 140, 25)$,

$$A \equiv \begin{bmatrix} 1 & 4 & 17 & 2 & 3 & 4 & 13 & 3 \\ 2 & 6 & 14 & 9 & 13 & 3 & 24 & 6 \end{bmatrix}, \quad \overline{b} = \begin{pmatrix} 28 \\ 38 \end{pmatrix} .$$

Penalties $qz \equiv (5,5)\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$, $r = 0$ were selected, giving the elastic resource problem:

$$\begin{aligned} V(\theta) = \underset{\substack{y=0,1 \\ z\geq 0}}{\text{Maximum}} \ \overset{8}{\underset{i=1}{\Sigma}} \ c_i y_i \ - \ \theta(5,5)\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\ \\ \text{subject to:} \ \ Ay - \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \leq \overline{b} \ . \end{aligned} \tag{11.15}$$

It was found that $\hat{z} = 0$ for $\theta = 2$ (hence, for all $\theta \geq 2$, $\hat{z} = 0$ and no net improvements to $\overline{v}(\overline{b})$ are possible); at $\theta = 1$, $\hat{z} = (0,3)$, affording an improvement of about 7% in the optimal value of (11.14); etc. Results are graphed in Figure 11.2 which shows $V(\theta)$, $\hat{z}(\theta)$, and $\overline{v}(\overline{b}+\hat{z}(\theta))$. (Since (11.14) is a maximization problem the usual senses of monotonicity are reversed.)

184
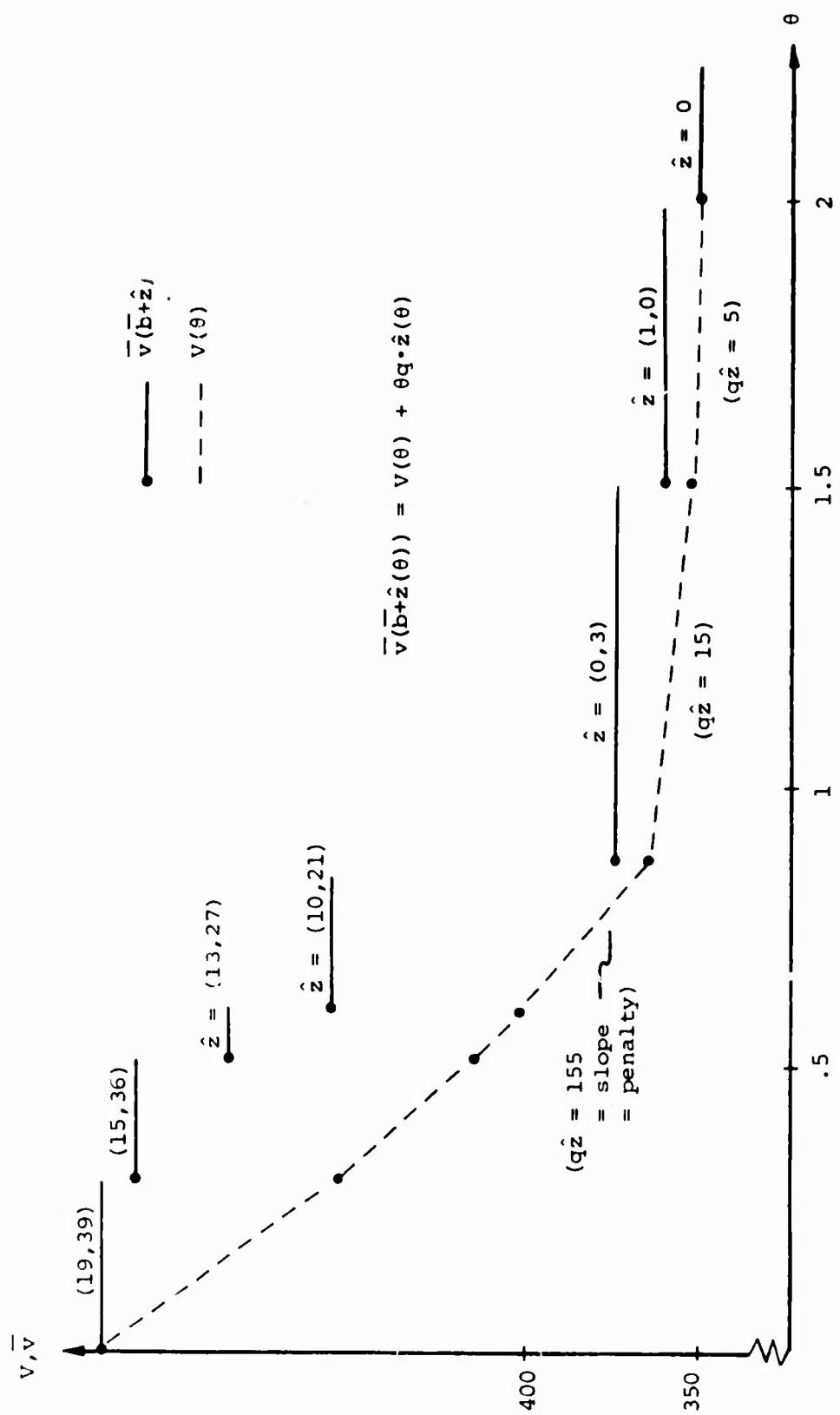
Figure 11.2. The Functions $V(\theta)$ and $\bar{v}(\bar{b}+\hat{z}(\theta))$ for Example (11.14)

$\hat{z} = 0$

$\hat{z} = (1,0)$
$(q\hat{z} = 5)$

$\hat{z} = (0,3)$
$(q\hat{z} = 15)$

$\hat{z} = (13,27)$

$\hat{z} = (10,21)$

$\bar{v}(\bar{b}+\hat{z}(\theta)) = V(\theta) + \theta q \cdot \hat{z}(\theta)$

$\bar{v}(\bar{b}+\hat{z})$

$V(\theta)$

$(15,36)$

$(19,39)$

$(q\hat{z} = 155$
$= slope$
$= penalty)$

$V, \bar{v}$

400

350

.5   1   1.5   2

θ

185

Several penalties were selected to determine the effect on the set of alternative solutions. These included $q = (6,4)$, $(4,6)$, $(7,3)$, and $(3,7)$, all with $r = 0$. In each of these cases, deriving $V(\theta)$ over $\theta \geq 0$ produced the same set of perturbations $\hat{z}$ and hence the same alternative solutions.

In linear problems such as example (11.15) the function $V(\theta)$ is piecewise linear; corresponding to each piece is an alternative solution $(\hat{x}, \hat{y}, \hat{z})$ to (P) with optimal value $\overline{v}(\overline{b} + \hat{z})$. That is, $\hat{z}(\theta)$ and $\overline{v}(\overline{b} + \hat{z}(\theta))$ are piecewise constant.

The set S of solutions $\hat{z}$ generated by deriving $V(\theta)$ is efficient in the sense that if $\hat{z} \in S$, then $\overline{v}(\overline{b} + \hat{z}) < \overline{v}(\overline{b} + z)$ and $p(\hat{z}) \leq p(z)$ imply that $z \notin S$.

To compare alternative solutions to (P) the decision-maker essentially must perform for each alternative a tradeoff between the amount of improvement Δv provided in $\overline{v}$ and the increase $\hat{z}$ in $\overline{b}$ required to produce that improvement. Whether or not a set S of alternative solutions includes the "most preferred" alternative depends on the decision-maker's tradeoff function. If this unknown function were linear in $f(\hat{x}, \hat{y})$ and $p(\hat{z})$, then, of course, using a parametric penalty $\theta p(z)$ in the elastic resource model (Q) and deriving $V(\theta)$ over a reasonable range of values θ will assuredly produce the decision-maker's most preferred alternative.

11.4   Summary and Conclusions

Relatively inexpensive bounding tests can be made to determine if substantial improvements in $\overline{v}(\overline{b})$ are possible (or impossible) by allowing a perturbation to the right-side $\overline{b}$. If these tests imply that

improved solutions are possible, then any one of several methods can be used to discover alternative solutions. Two particularly attractive approaches are rounding and use of elastic resource formulations.

Optimality of a rounded solution at a perturbed right-side is guaranteed in certain classes of MIP problems (Th. 11.1). In the general case, the degree of suboptimality in the MIP of a rounded solution is easily computed (cf. Th. 10.4). "Optimal" rounded solutions can be derived via relatively small 0-1 integer programs.

An elastic resource approach gives recourse to the resource constraints by allowing them to be violated for a price. In terms of a given cost criterion $p(\Delta b)$, this approach provides a solution to the MIP problem that gives the maximum net improvement in optimal value. A collection of alternative solutions can be generated by varying the penalty cost linearly in terms of a scalar parameter θ. This parametrization effectively converts the right-side analysis problem to a problem parametric in the objective function, for which straightforward parametric procedures exist.

The attractiveness of a rounding approach is that alternative solutions to the MIP problem are provided very cheaply by solving a continuous-variable relaxation (and possibly a small 0-1 integer problem); the elastic resource approach involves solving another MIP problem. On the other hand, a rounding approach cannot involve the decision-maker's "costs" assigned to a perturbation Δb and it is possible that although rounding produces a solution to the MIP which grants an improvement in optimal value, this improvement may be overshadowed by the decision-maker's costs assigned to the required perturbation Δb. The elastic resource approach permits one to consider explicitly in the model an

estimate of the decision-maker's costs.

An elastic resource model specializes easily to the case where only right-side changes in a direction b' are of incerest. Then z is a scalar variable and the perturbation to $\bar{b}$ is given by $\Delta b = zb'$.

BIBLIOGRAPHY - PART II

1. Everett, H., "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," _Operations Research_, Vol. 11 (1963).

2. Gale, D., _The Theory of Linear Economic Models_, McGraw-Hill, New York, 1960.

3. Garfinkel, R.S., and G.L. Nemhauser, _Integer Programming_, John Wiley, New York, 1972.

4. Geoffrion, A.M., "Elements of Large-Scale Mathematical Programming," _Management Science_, Vol. 16, No. 11 (July 1970).

5. _____, "Duality in Nonlinear Programming: A Simplified Applications-Oriented Development," _SIAM Review_, Vol. 13, No. 1, January 1971.

6. _____, "Lagrangean Relaxation for Integer Programming," _Mathematical Programming Study 2_ (1974).

7. Hespel, P.P., "Reimputing the Dual Variables in Integer Programming," Discussion Paper No. 6707, Center for Operations Research and Econometrics, Universite Catholique de Louvain, Heverlee, Belgium, May 1967.

8. Jensen, R.E., "Sensitivity Analysis and Integer Linear Programming," _The Accounting Review_ (July 1968).

9. Lasdon, L.S., _Optimization Theory for Large Systems_, Macmillan, New York, 1970.

10. Luenberger, D.G., _Introduction to Linear and Nonlinear Programming_, Addison-Wesley, Reading, Mass., 1973.

11. Manne, A.S. (ed.), _Investments for Capacity Expansion: Size, Location, and Time-Phasing_, The MIT Press, Cambridge, Mass., 1967.

12. Nauss, R.M., "Parametric Integer Programming," Working Paper No. 226, Western Management Science Institute, UCLA, January 1975.

13. Noltemeier, H., "Sensitivitätsanalyse bei diskreten linearen Optimierungsproblemen," _Lecture Notes in Operations Research and Mathematical Systems_, #30, M. Beckman and H.P. Kunzi (eds.), Springer-Verlag, New York, 1970.

14. Rappaport, A., "Sensitivity Analysis in Decision Making," _The Accounting Review_ (July 1967).

15. Roode, J.D., _Generalized Lagrangian Functions in Mathematical Programming_, Rotterdam, 1968.

16. Roodman, G.M., "Postoptimality Analysis in Zero-One Programming by Implicit Enumeration," *Naval Research Logistics Quarterly*, Vol. 19 (1972).

17. _____, "Postoptimality Analysis in Integer Programming by Implicit Enumeration: The Mixed Integer Case," Working Paper, Amos Tuck School of Business Administration, Dartmouth College, October 1973.

18. Schrage, L., "Implicit Representation of Variable Upper Bounds in Linear Programming," Working Paper No. 217, Western Management Science Institute, UCLA, July 1974.

19. Weingartner, H.M., *Mathematical Programming and the Analysis of Capital Budgeting Problems*, Prentice-Hall, Englewood Cliffs, New Jersey, 1963.

PART III

CHAPTER 12

CONCLUDING REMARKS AND TOPICS OF FURTHER STUDY

Sensitivity analysis in mixed integer programming is a new and much needed area of study that will undoubtedly assume greater significance as MIP solution methods advance. It is acknowledged that this treatise on the subject is a mere beginning. It is hoped that our comprehensive theory of continuity provides greater understanding of the causes and ways of dealing with the discontinuous behavior that is typical of MIP problems. We consider the treatment of sensitivity theory and analysis to be both very general and practical; to be fully exploited, various specializations and computational experience with special classes of problems will be required as a next step.

In developing and specializing analysis techniques to problem classes it is important to consider the type of code required for implementation. A requirement for a very general MIP code to analyze a special type of problem, that is usually solved with a special code, can be a deterrent to conducting analysis. In facility location problems, for instance, it is easy to show that the elastic resource approach can be implemented by simply adding a dummy facility and a dummy demand zone to the problem, still permitting the usual transportation type subproblem.

Numerous computational studies are called for. These include studies to: (a) evaluate and compare merits of the various analysis methods in terms of the information produced, (b) determine computational efficiencies and requirements, and (c) gain a better

191

understanding of the behavior of special classes of problems. It is intuitive that an elastic resource relaxation (e.g., (Q), §11.3) should be algorithmically easier to solve because the z variables make it easier for a feasible solution generator to achieve feasible rounded solutions. Of course, computational experience is needed to bear this out. It is also intuitive that rounding will produce attractive alternative solutions for problems with a large number of variables, especially when the number of fractional variables is relatively small.

There are other important areas of further research. One of these, perhaps the most important, is the design of computer runs. Run design involves the definition and sequencing of problems to be solved that will yield the desired information at the least computational expense.

The basis for run design is a priori knowledge of how the optimal solution (x^*, y^*) behaves with respect to the data. This type of knowledge can suggest not only an efficient sequencing of runs but how the most information per run can be obtained. As an example, consider the capacitated facility location problem (cf. §4.2,. If a run is repeated with all capacities and demands scaled by a factor θ, the result can be used to predict by inspection the consequences of repeating the original run with the fixed costs (y coefficients) scaled by a factor of $1/\theta$, and conversely. There are also other reciprocal relationships.

Another interesting area for future research concerns the decision-maker's unknown preferences for alternative solutions. Since a collection of alternative solutions to an MIP problem is generated for the purpose of providing the decision-maker with a choice that best satisfies his preferences, one naturally must be concerned with the possibility that this set excludes the most preferred alternative. If the

192

decision-maker's preference is represented by some (unknown) function $h[\bar{v}, p(\Delta b)]$ where $\bar{v}$ is the improved optimal value and Δb the associated perturbation to the RS $\bar{b}$, then the preferred alternative is given by

$$
(H) \quad
\begin{cases}
\text{Minimize } h[f(x,y), p(z)] \\
\quad x,y,z \\[2mm]
\text{subject to: } G(x,y) - z \leq \bar{b} \\[2mm]
\qquad\qquad (x,y) \in U \ .
\end{cases}
$$

Problems of this type with convex domains are studied by Geoffrion, Dyer and Feinberg [1] and others; these developments, however, do not apply to problems with integer restrictions. Conditions on h, f, p, etc., and approaches to generating alternative solutions are desired which assure that a set of alternatives includes a solution to (H).

BIBLIOGRAPHY - PART III

1. Geoffrion, A.M., J.S. Dyer, and A. Feinberg, "An Interactive Approach for Multi-Criterion Optimization, with an Application to the Operation of an Academic Department," *Management Science*, Vol. 19, No. 4 (December 1972), Part I.